

# A Behavioral Framework for Data-Driven Modeling of Nonlinear Systems in Vector-Valued Reproducing Kernel Hilbert Spaces

Boya Hou<sup>1</sup> and Maxim Raginsky<sup>2</sup>

**Abstract**—We generalize Jan Willems’ behavioral approach to a class of discrete-time nonlinear systems in a vector-valued reproducing kernel Hilbert space (RKHS). Apart from linear time-invariant systems, this class covers nonlinear systems modeled by Volterra series and their autoregressive variants, as well as systems admitting Hammerstein-type state-space realizations. We apply the proposed framework to the problem of data-driven modeling of such systems, i.e., when simulation or control objectives for an unknown system are carried out without an explicit system identification step. To that end, we link the behavioral approach to two data-driven modeling methods in a vector-valued RKHS: (1) minimum-norm interpolation and (2) subspace identification.

## I. INTRODUCTION

The mathematical study of dynamical systems seeks to describe the evolution of a given system over time under its governing laws, initial conditions, and inputs (if any). This evolution can be studied internally through state-space methods or externally via the system’s behavior. Pioneered by Jan Willems in a series of papers starting with [1] (see [2] for an overview), the behavioral approach identifies dynamical systems with sets of their trajectories. For linear time-invariant systems, the cornerstone of this framework is the so-called *fundamental lemma* [3], which asserts that the set of all trajectories of a controllable linear time-invariant system over a finite time horizon can be reconstructed from a single trajectory driven by a persistently exciting input. This result has become central to recent advances of data-driven control; see, e.g., [4, 5, 6, 7, 8] and references therein.

The philosophy underlying these data-driven approaches, as clearly articulated in the paper by Markovsky and Rapisarda [5], is that simulation or control should be carried out without an explicit system identification step. In the context of simulation of linear systems, for example, this principle is supported by the fact that, given an initial condition and a subsequent input of interest, one can combine it with previous measurements of the system behavior to predict the resulting output without identifying the system transfer function, impulse response, or state-space model. Rather, simulation can be viewed as a “missing data” problem, to be solved directly using the available trajectories.

While this viewpoint has led to a rich and mature literature for linear systems, extending it to nonlinear systems is substantially more challenging, as the trajectory sets no

longer form linear subspaces. Several works have explored nonlinear analogues of the fundamental lemma for structured nonlinear systems. Examples include bilinear systems [9], second-order Volterra systems [10], flat nonlinear systems [11], Hammerstein and Wiener systems [12], and a Koopman-type linear embedding posited in [13]. A promising direction is to use a reproducing kernel Hilbert space (RKHS) to lift the trajectories of nonlinear systems into a (possibly infinite-dimensional) feature space. In this vein, Huang, Lygeros, and Dörfler [14] investigated kernelized data-enabled predictive control, where the resulting models are regression-based approximations rather than behavioral characterizations. More recently, Molodchyk and Faulwasser [15] established the link between kernel regression and Willems’ fundamental lemma, and showed that several existing nonlinear extensions of the fundamental lemma, such as those for Hammerstein and flat systems, can be interpreted as instances of kernel regression with specific choices of kernel when the induced RKHS is finite-dimensional.

In this paper, we extend the behavioral framework beyond the linear settings for characterizing the behavior of nonlinear systems in RKHS via minimum-norm interpolation and subspace identification repurposed for nonlinear systems. We reinterpret both minimum-norm interpolation and subspace identification through the behavioral lens, and clarify the role of various structural assumptions on the system, the offline and online data, and the space of functions used to represent the nonlinear aspects of the system. A unifying theme across these two methods is that, like in the modern data-driven schemes and in the spirit of the fundamental lemma, predictors and the desired “online” trajectories can be expressed in terms of the kernelized “offline” examples. When working in an RKHS, as noted by Molodchyk and Faulwasser [15], this structure further mirrors the representer theorem [16]: the solution can be written as a finite linear combination of kernel evaluations along the observed trajectories.

The remainder of this paper is organized as follows. In Section II, we recall the fundamental lemma of Willems for linear time-invariant systems and introduce basic concepts regarding scalar-valued and vector-valued RKHSs. In Section III, we introduce a class of nonlinear systems whose nonlinearity is encoded by an element in the vector-valued RKHS of functions on the set of finite-length sequences of inputs. In Section IV, we establish a *Behavior Representer Theorem* for nonlinear systems using minimum-norm interpolation and contrast it with the fundamental lemma. Compared with the work of Molodchyk and Faulwasser [15], we consider minimum-norm interpolation instead of (regularized) least-squares regression in a vector-valued RKHS. We further establish an error estimate that holds with *equality* as in Lemma 4, rather than the inequality in [15,

This work was supported in part by the NSF under award CCF-2348624 (“Towards a control framework for neural generative modeling”).

<sup>1</sup> Boya Hou is with the Carl R. Woese Institute for Genomic Biology and the Coordinated Science Laboratory, University of Illinois Urbana-Champaign, Urbana, IL, 61801 USA boyahou2@illinois.edu

<sup>2</sup> Maxim Raginsky is with Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois Urbana-Champaign, Urbana, IL, 61801 USA maxim@illinois.edu

Lemma 2]. In Section V, we present a subspace identification framework for nonlinear systems in an RKHS and provide a fundamental lemma characterization of finite-length trajectories with kernelized input vectors and outputs.

## II. PRELIMINARIES

### A. Notation and definitions

We will make use of the following notation and definitions throughout the paper. We will use  $\mathbb{Z}_+$  to denote the set of nonnegative integers. The Moore–Penrose pseudoinverse of a matrix  $A$  will be denoted by  $A^\dagger$ . Given the matrices  $A \in \mathbb{R}^{p \times k}$ ,  $B \in \mathbb{R}^{q \times k}$ , and  $C \in \mathbb{R}^{r \times k}$ , the *oblique projection* of the row space of  $A$  on the row space of  $C$  along the row space of  $B$  is defined as

$$A/C \underset{B}{:=} A \begin{bmatrix} C^r & B^r \end{bmatrix} \left( \begin{bmatrix} CC^r & CB^r \\ BC^r & BB^r \end{bmatrix}^\dagger \right) \underset{\text{first } r \text{ columns}}{C} \quad C$$

(see [17, Section 1.4.2]). Given a finite sequence of vectors  $w_{0:T-1} = \{w_t\}_{t=0}^{T-1}$  in  $\mathbb{R}^q$ , the *Hankel matrix of depth  $L$*  is the  $(Lq) \times (T-L+1)$  matrix given by

$$H_L(w_{0:T-1}) := \begin{bmatrix} w_0 & w_1 & \dots & w_{T-L} \\ w_1 & w_2 & \dots & w_{T-L+1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{L-1} & w_L & \dots & w_{T-1} \end{bmatrix}.$$

We say that  $w_{0:T-1}$  is *persistently exciting (PE) of order  $L$*  if the Hankel matrix  $H_L(w_{0:T-1})$  has full row rank. Given a signal (discrete-time vector-valued sequence)  $\{w_t\}_{t \in \mathbb{Z}_+}$ ,  $\sigma$  is the backward shift operator defined by  $(\sigma w)_t := w_{t+1}$ . Additional notation will be introduced as needed.

### B. A review of the fundamental lemma

In the behavioral framework, a linear time-invariant system with  $q$  variables is identified with a linear subspace  $\mathcal{B}$  of the space of all one-sided sequences of  $q$ -dimensional real vectors denoted as  $(\mathbb{R}^q)^{\mathbb{Z}_+}$  which is shift-invariant, i.e.,  $\sigma \mathcal{B} \subseteq \mathcal{B}$ . One can work with various equivalent *representations* of  $\mathcal{B}$ , such as autoregressive, input/output, or input/state/output representations [1]. Notions like controllability or observability can be defined solely in terms of set-theoretic properties of  $\mathcal{B}$ , without referring to a particular representation [2].

One of these structural properties pertains to the partition of the variables of  $\mathcal{B}$  into inputs and outputs. Without getting too much into technical details, the idea is that, up to a permutation of coordinates, each trajectory  $w \in \mathcal{B}$  can be partitioned into an input trajectory  $u : \mathbb{Z}_+ \rightarrow \mathbb{R}^m$  and an output trajectory  $y : \mathbb{Z}_+ \rightarrow \mathbb{R}^p$  as  $w = \begin{bmatrix} u \\ y \end{bmatrix}$ , such that the input is free in the sense that

$$\left\{ u : \mathbb{Z}_+ \rightarrow \mathbb{R}^m \mid \begin{bmatrix} u \\ y \end{bmatrix} \in \mathcal{B} \text{ for some } y : \mathbb{Z}_+ \rightarrow \mathbb{R}^p \right\}, \\ \simeq (\mathbb{R}^m)^{\mathbb{Z}_+}$$

and the output is determined by the input, subject to the system laws and the initial condition. It can be shown that the number

of inputs  $m$ , the number of outputs  $p$ , and the dimension  $n$  of any minimal state space realization of  $\mathcal{B}$  are system invariants that depend only on  $\mathcal{B}$  and not on the particular representation of  $\mathcal{B}$  [1].

Let  $\mathcal{B}$  be a controllable linear time-invariant system with  $m$  inputs,  $p$  outputs, and minimum state dimension  $n$ . The main result of [3], now commonly referred to as the *fundamental lemma*, is as follows:

**Lemma 1.** *For each  $t = 1, 2, \dots$ , let  $\mathcal{B}|_t$  denote the restriction of  $\mathcal{B}$  to times  $s \in \{0, \dots, t-1\}$ . Let a trajectory  $w_{0:T-1}^d \in \mathcal{B}|_T$  be given, such that the input part of  $w_{0:T-1}^d$  is persistently exciting of order  $L+n$ . Then  $\mathcal{B}|_L$  is equal to the column space of the Hankel matrix  $H_L(w_{0:T-1}^d)$ .*

The main message of Lemma 1 is that the length- $L$  behavior  $\mathcal{B}|_L$  can be reconstructed, exactly and in a representation-independent manner, from a single input/output trajectory of length  $T \geq L + (L+n)m - 1$ , provided the input is persistently exciting and the system is controllable. This makes the fundamental lemma a key ingredient in data-driven approaches to system simulation and control.

### C. Reproducing Kernel Hilbert Spaces

In this paper, we make extensive use of vector-valued reproducing kernel Hilbert spaces. We start by describing the more familiar definition of a scalar-valued RKHS; we refer interested readers to [18] for details. Let  $X$  be a set. A Hilbert space<sup>1</sup>  $\mathcal{H}$  with inner product  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  is an RKHS on  $X$  if its elements are functions from  $X$  to  $\mathbb{R}$  and if for each  $x \in X$  there exists a positive constant  $C_x$ , such that  $|f(x)| \leq C_x \|f\|_{\mathcal{H}}$  for all  $f \in \mathcal{H}$ . In other words,  $\mathcal{H}$  is an RKHS on  $X$  if the evaluation functional  $\delta_x(f) := f(x)$  is bounded (hence continuous) for all  $x \in X$ . To each RKHS, we can associate a *reproducing kernel*, i.e., a mapping  $\kappa : X \times X \rightarrow \mathbb{R}$  that satisfies the positivity condition

$$\sum_{i,j=1}^n c_i c_j \kappa(x_i, x_j) \geq 0$$

for all  $n \in \mathbb{N}$ ,  $c_1, \dots, c_n \in \mathbb{R}$ , and  $x_1, \dots, x_n \in X$ , such that  $\kappa(\cdot, x)$  is an element of  $\mathcal{H}$  for each  $x \in X$  and the following property (called the *reproducing kernel property*) holds:

$$f(x) = \langle f, \kappa(\cdot, x) \rangle_{\mathcal{H}}, \quad \text{for all } f \in \mathcal{H}, x \in X.$$

Moreover, the set  $\{\kappa(\cdot, x) : x \in X\}$  is total in  $\mathcal{H}$  (i.e., its linear span is dense in  $\mathcal{H}$ ). The map  $\phi : X \rightarrow \mathcal{H}$  defined by  $\phi(x) := \kappa(\cdot, x)$  is called the *canonical feature map*. By the Moore–Aronszajn theorem [19], the reproducing kernel  $\kappa$  specifies  $\mathcal{H}$  uniquely up to linear isomorphism.

We now describe the vector-valued generalization of the above definition [20]. Let  $\mathcal{K}$  be a Hilbert space, and let  $\mathcal{L}(\mathcal{K})$  denote the Hilbert space of bounded linear operators on  $\mathcal{K}$  with the Hilbert–Schmidt norm. Then we say that  $\mathcal{H}$  is a  *$\mathcal{K}$ -valued RKHS on  $X$*  if its elements are functions from  $X$  to  $\mathcal{K}$  and if the evaluation map  $\delta_x(f) := f(x)$  from  $\mathcal{H}$  to  $\mathcal{K}$  is bounded

<sup>1</sup>Unless indicated otherwise, all Hilbert spaces in this paper are assumed to be defined over the reals.

(hence continuous) for all  $x \in X$ , i.e., there exists a positive constant  $C_x$  such that  $\|f(x)\|_{\mathcal{K}} \leq C_x \|f\|_{\mathcal{H}}$ . The vector-valued analogue of the reproducing kernel is an *operator-valued* map  $\kappa : X \times X \rightarrow \mathcal{L}(\mathcal{K})$  of the positive type, i.e., such that the inequality

$$\sum_{i,j=1}^n c_i c_j \langle \kappa(x_i, x_j) v, v \rangle_{\mathcal{K}} \geq 0$$

holds for all  $n \in \mathbb{N}$ ,  $c_1, \dots, c_n \in \mathbb{R}$ ,  $x_1, \dots, x_n \in X$ , and  $v \in \mathcal{K}$ . The reproducing kernel property then takes the form

$$\langle f(x), v \rangle_{\mathcal{K}} = \langle f, \kappa(\cdot, x) v \rangle_{\mathcal{H}}, \text{ for all } f \in \mathcal{H}, x \in X, v \in \mathcal{K}$$

where, for each  $x \in X$ ,  $\kappa(\cdot, x)$  is an element of the linear space  $\mathcal{L}(\mathcal{K}, \mathcal{H})$  of bounded operators from  $\mathcal{K}$  to  $\mathcal{H}$ . The map  $x \mapsto \kappa(\cdot, x)$  from  $X$  into  $\mathcal{L}(\mathcal{K}, \mathcal{H})$  is the (operator-valued) canonical feature map.

A basic example of a vector-valued RKHS which will be used frequently in the sequel is  $\mathcal{H} = \mathcal{L}(\mathcal{V}, \mathcal{W})$ , where  $\mathcal{V}$  and  $\mathcal{W}$  are finite-dimensional inner-product spaces and where we equip  $\mathcal{H}$  with the Hilbert–Schmidt inner product  $\langle A, B \rangle_{\mathcal{H}} = \text{Tr}(A^* B)$ . For the sake of completeness, we present the straightforward proof of the following lemma in Appendix A.

**Lemma 2.**  *$\mathcal{H}$  is a  $\mathcal{W}$ -valued reproducing kernel Hilbert space on  $\mathcal{V}$  with the operator-valued reproducing kernel  $\kappa(v, v') = \langle v, v' \rangle_{\mathcal{V}} I_{\mathcal{W}}$ , where  $I_{\mathcal{W}}$  is the identity operator on  $\mathcal{W}$ .*

### III. NONLINEAR SYSTEMS IN A VECTOR-VALUED RKHS

We now introduce a class of discrete-time nonlinear systems, where the nonlinearity is represented by an element of a given vector-valued RKHS of functions on the set of finite-length sequences of inputs. We will use the following notation throughout:

- $\mathcal{U} := \mathbb{R}^m$  is the space of inputs;
- $\mathcal{Y} := \mathbb{R}^p$  is the space of outputs;
- $\mathcal{U} := \mathcal{U}^{L+1}$  is the space of length- $(L+1)$  input sequences, where  $L \in \mathbb{Z}_+$  is a fixed lag;
- $\mathcal{Y} := \mathcal{Y}^L$  is the space of length- $L$  output sequences;
- $\mathcal{Z} := \mathcal{U} \times \mathcal{Y}$ .

We introduce the system model in Section III-A and the associated behavioral constructs in Section III-B, and close by discussing several examples in Section III-C.

#### A. System model

Let  $\mathcal{H}_{\mathcal{U}}$  be a vector-valued RKHS of functions from  $\mathcal{U}$  into  $\mathcal{Y}$ , with the operator-valued kernel  $\kappa_{\mathcal{U}}$  of positive type.<sup>2</sup> We consider systems parametrized by  $(L+1)$ -tuples  $(A_1, \dots, A_L, g)$ , where  $A_1, \dots, A_L \in \mathcal{L}(\mathcal{Y})$  and  $g \in \mathcal{H}_{\mathcal{U}}$ . The input/output relation corresponding to  $(A_1, \dots, A_L, g)$  is given by

$$y_{t+L} + \sum_{k=1}^L A_k y_{t+L-k} = g(u_{t:t+L}), \quad t \in \mathbb{Z}_+ \quad (1)$$

<sup>2</sup>Here and elsewhere, all finite-dimensional vector spaces are automatically treated as Hilbert spaces with the usual Euclidean inner product.

where  $u_{t:t+L}$  is the restriction of the input sequence  $u \in \mathcal{U}^{\mathbb{Z}_+}$  to the set  $\{t, t+1, \dots, t+L\}$ . By the reproducing kernel property, we can also write

$$g(u_{t:t+L}) = \kappa_{\mathcal{U}}(\cdot, u_{t:t+L})^* g,$$

where  $\kappa_{\mathcal{U}}(\cdot, u_{t:t+L})^* \in \mathcal{L}(\mathcal{H}_{\mathcal{U}}, \mathcal{Y})$  is the adjoint of the canonical feature map  $\kappa_{\mathcal{U}}(\cdot, u_{t:t+L}) \in \mathcal{L}(\mathcal{Y}, \mathcal{H}_{\mathcal{U}})$ .

It is also convenient to cast (1) in an alternative nonlinear regression form. For each  $t$ , let  $y_{t+}$  denote the output  $y_{t+L}$  at time  $t+L$  and define the *regression vectors*

$$z_t := [u_t^T, \dots, u_{t+L}^T, y_t^T, \dots, y_{t+L-1}^T]^T \in \mathcal{Z} \quad (2)$$

in analogy to the corresponding construct for linear systems [21, 22]. As we now show, we can introduce a  $\mathcal{Y}$ -valued RKHS  $\mathcal{H}_{\mathcal{Z}}$  of functions from  $\mathcal{Z}$  into  $\mathcal{Y}$ , such that the autoregressive nonlinear model of (1) can be represented as

$$y_{t+} = f(z_t) \quad (3)$$

for some  $f \in \mathcal{H}_{\mathcal{Z}}$  that depends on  $(A_1, \dots, A_L, g)$ .

To that end, let us first express (1) as

$$\begin{aligned} y_{t+L} &= A^- y_{t:t+L-1} + g(u_{t:t+L}) \\ &=: f(u_{t:t+L}, y_{t:t+L-1}), \end{aligned} \quad (4)$$

where  $A^- \in \mathcal{L}(\mathcal{Y}, \mathcal{Y})$  is the linear operator defined by

$$A^- \bar{y} := -[A_1 \ A_2 \ \dots \ A_L] \bar{y}, \quad \bar{y} \in \mathcal{Y}.$$

By Lemma 2,  $\mathcal{H}_{\mathcal{Y}} = \mathcal{L}(\mathcal{Y}, \mathcal{Y})$  is a  $\mathcal{Y}$ -valued RKHS on  $\mathcal{Y}$  with the reproducing kernel  $\kappa_{\mathcal{Y}}(\bar{y}, \bar{y}') = (\bar{y}'^T \bar{y}) I_{\mathcal{Y}}$ . Consider the direct-sum Hilbert space  $\mathcal{H}_{\mathcal{Z}} = \mathcal{H}_{\mathcal{U}} \oplus \mathcal{H}_{\mathcal{Y}}$ , whose elements  $f : \mathcal{Z} \rightarrow \mathcal{Y}$  are pairs  $(g, h)$  where  $g \in \mathcal{H}_{\mathcal{U}}$ ,  $h \in \mathcal{H}_{\mathcal{Y}}$ , and for each  $z = (\bar{u}, \bar{y}) \in \mathcal{Z}$ ,  $f(z) = g(\bar{u}) + h(\bar{y})$ . The inner product in  $\mathcal{H}_{\mathcal{Z}}$  is

$$\langle f_1, f_2 \rangle_{\mathcal{H}_{\mathcal{Z}}} = \langle g_1, g_2 \rangle_{\mathcal{H}_{\mathcal{U}}} + \langle h_1, h_2 \rangle_{\mathcal{H}_{\mathcal{Y}}},$$

for  $f_1 = (g_1, h_1)$ ,  $f_2 = (g_2, h_2)$ . Since  $\mathcal{H}_{\mathcal{U}}$ ,  $\mathcal{H}_{\mathcal{Y}}$  are  $\mathcal{Y}$ -valued RKHSs with operator-valued kernels  $\kappa_{\mathcal{U}}$  and  $\kappa_{\mathcal{Y}}$ ,  $\mathcal{H}_{\mathcal{Z}}$  is also an RKHS with operator-valued kernel defined by

$$\begin{aligned} \kappa_{\mathcal{Z}}(z_1, z_2) &= \kappa_{\mathcal{U}}(\bar{u}_1, \bar{u}_2) + \kappa_{\mathcal{Y}}(\bar{y}_1, \bar{y}_2), \\ z_1 &= (\bar{u}_1, \bar{y}_1), z_2 = (\bar{u}_2, \bar{y}_2) \in \mathcal{Z}. \end{aligned}$$

This can be proved as follows: for any  $f \in \mathcal{H}_{\mathcal{Z}}$  and  $v \in \mathcal{Y}$ , we have

$$\begin{aligned} \langle f(z), v \rangle_{\mathcal{Y}} &= \langle g(z), v \rangle_{\mathcal{Y}} + \langle h(z), v \rangle_{\mathcal{Y}} \\ &= \langle g, \kappa_{\mathcal{U}}(\cdot, \bar{u}) v \rangle_{\mathcal{H}_{\mathcal{U}}} + \langle h, \kappa_{\mathcal{Y}}(\cdot, \bar{y}) v \rangle_{\mathcal{H}_{\mathcal{Y}}} \\ &= \langle f, (\kappa_{\mathcal{U}}(\cdot, \bar{u}) + \kappa_{\mathcal{Y}}(\cdot, \bar{y})) v \rangle_{\mathcal{H}_{\mathcal{Z}}}, \end{aligned}$$

where the second line follows from the reproducing kernel property. In particular, (3) holds with the function  $f$  defined in (4).

## B. Behaviors

In this section, we introduce several behavioral constructs related to the system model of Section III-A.

Given  $(A_1, \dots, A_L, g)$ , define the operator  $P(\sigma) := A_0 \sigma^L + \sum_{k=1}^L A_k \sigma^{L-k}$  (with  $A_0 = I_{\mathcal{Y}}$ ), where  $\sigma$  is the shift operator acting on output sequences. Then (1) is equivalent to

$$(P(\sigma)y)(t) = \kappa_{\mathcal{U}}(\cdot, u_{t:t+L})^* g, \quad t \in \mathbb{Z}_+ \quad (5)$$

and we can define the behavior of  $(A_1, \dots, A_L, g)$  as the following subset of the space of all input/output sequences:

$$\begin{aligned} & \mathcal{B}(A_1, \dots, A_L, g) \\ & \equiv \mathcal{B}(P, g) \\ & := \left\{ \begin{bmatrix} u \\ y \end{bmatrix} \in (\mathcal{U} \times \mathcal{Y})^{\mathbb{Z}_+} \mid (5) \text{ holds for all } t \in \mathbb{Z}_+ \right\}. \end{aligned} \quad (6)$$

For  $t \in \mathbb{Z}_+$ , the restriction of the behavior  $\mathcal{B}(P, g)$  to the finite time interval  $\{t, \dots, t+L\}$  is defined by

$$\begin{aligned} \mathcal{B}(P, g)|_{t:t+L} &= \left\{ \begin{bmatrix} u_{t:t+L} \\ y_{t:t+L} \end{bmatrix} \mid \exists \begin{bmatrix} \bar{u} \\ \bar{y} \end{bmatrix} \in \mathcal{B}(P, g) \right. \\ & \left. \text{such that } \begin{bmatrix} \bar{u}_{t:t+L} \\ \bar{y}_{t:t+L} \end{bmatrix} = \begin{bmatrix} u_{t:t+L} \\ y_{t:t+L} \end{bmatrix} \right\}. \end{aligned}$$

When  $t = 0$ , we will use  $\mathcal{B}(P, g)|_L$  as shorthand for  $\mathcal{B}(P, g)|_{t:t+L}$ .

Next, we turn to the nonlinear regression form in (3). Recall that the identity

$$f(z) = \kappa_{\mathcal{Z}}(\cdot, z)^* f$$

holds by the reproducing kernel property. Define the operator  $R : (\mathcal{U} \times \mathcal{Y})^{\mathbb{Z}_+} \rightarrow \mathcal{L}(\mathcal{H}_{\mathcal{Z}}, \mathcal{Y}) \times \mathcal{Y}$  by

$$R \begin{bmatrix} u \\ y \end{bmatrix} := (\kappa_{\mathcal{Z}}(\cdot, z_0)^*, y_{0+}),$$

where

$$z_0 = [u_0^T, \dots, u_L^T, y_0^T, \dots, y_{L-1}^T]^T, \quad y_{0+} = y_L \quad (7)$$

is the regression vector at time  $t = 0$  [cf. Eq. (2)] and the output at time  $t = L$ . We say that  $(H, y) \in \mathcal{L}(\mathcal{H}_{\mathcal{Z}}, \mathcal{Y}) \times \mathcal{Y}$  is an input-output (i/o) pair of the system (3) if  $Hf = y$ . This leads naturally to the following behavioral description:

$$\begin{aligned} \mathcal{B}(f) &:= \left\{ \begin{bmatrix} u \\ y \end{bmatrix} \in (\mathcal{U} \times \mathcal{Y})^{\mathbb{Z}_+} \mid R\sigma^t \begin{bmatrix} u \\ y \end{bmatrix} \right. \\ & \left. \text{is an i/o pair of (3) for each } t \in \mathbb{Z}_+ \right\}. \end{aligned}$$

When  $(A_1, \dots, A_L, g)$  and  $f$  are related via (3), it is easily verified that

$$\mathcal{B}(A_1, \dots, A_L, g)|_L = \left\{ \begin{bmatrix} u_{0:L} \\ y_{0:L} \end{bmatrix} \mid \kappa_{\mathcal{Z}}(\cdot, z_0)^* f = y_{0+} \right\}, \quad (8)$$

where  $z_0$  and  $y_{0+}$  are defined in (7).

## C. Examples

1) *LTI Systems*: Let  $\mathcal{H}_{\mathcal{U}}$  consist of all linear operators from  $\mathcal{U}$  into  $\mathcal{Y}$ , i.e.,  $\mathcal{H}_{\mathcal{U}} = \mathcal{L}(\mathcal{U}, \mathcal{Y})$ . By Lemma 2, this is a  $\mathcal{Y}$ -valued RKHS on  $\mathcal{U}$ . Any  $g \in \mathcal{H}_{\mathcal{U}}$  can be represented as  $g(u_{0:L}) = B_L u_0 + \dots + B_0 u_L$  for some linear operators  $B_0, \dots, B_L \in \mathcal{L}(\mathcal{U}, \mathcal{Y})$ . Then Eq. (1) describes a linear autoregressive model of order  $L$ :

$$y_{t+L} + \sum_{k=1}^L A_k y_{t+L-k} = \sum_{l=0}^L B_l u_{t+L-l}.$$

The operator-valued kernel map is given by

$$\kappa_{\mathcal{U}}(u_{0:L}, v_{0:L}) = \langle u_{0:L}, v_{0:L} \rangle_{\mathcal{U}} I_{\mathcal{Y}},$$

where

$$\langle u_{0:L}, v_{0:L} \rangle_{\mathcal{U}} = \sum_{l=0}^L u_l^T v_l$$

is the  $l^2$  inner product on  $\mathcal{U}$  viewed as a direct sum of  $L+1$  Hilbert spaces  $\mathcal{U} = \mathbb{R}^m$  with the Euclidean inner product  $\langle u, v \rangle = u^T v$ .

2) *Volterra series*: When  $A_k = 0$  for all  $k = 1, \dots, L$ , the nonlinear autoregressive model in (1) reduces to

$$y_{t+L} = g(u_{t:t+L}), \quad \forall t \in \mathbb{N}. \quad (9)$$

This system has finite memory of length  $L+1$ , since the output at each time  $t \geq L$  is determined by the inputs in the finite time window  $\{t, t-1, \dots, t-L\}$  of length  $L+1$ . Systems of this type are often represented using Volterra series. De Figueiredo and Dwyer [23] used the formalism of weighted Fock spaces to connect Volterra series representation to the RKHS framework.

Let us assume for simplicity that  $p = 1$  (i.e., the outputs  $y_t$  are scalar). Let  $\rho = (\rho_k)_{k \geq 0}$  be a sequence of positive reals, such that the infinite series

$$q(\lambda) = \sum_{k=0}^{\infty} \frac{1}{k! \rho_k} \lambda^k$$

converges for all  $\lambda \in \mathbb{R}$ , and let  $\mathcal{H}_{\mathcal{U}}$  denote the RKHS of functions from  $\mathcal{U}$  into  $\mathbb{R}$  with the reproducing kernel

$$\kappa_{\mathcal{U}}(u_{0:L}, v_{0:L}) := q(\langle u_{0:L}, v_{0:L} \rangle_{\mathcal{U}}).$$

Let  $h_0$  and  $(h_k(i_1, \dots, i_k) : k \in \mathbb{N}, i_1, \dots, i_k \in \{0, \dots, L\})$  be a collection of real coefficients satisfying the condition

$$\sum_{k=0}^{\infty} \frac{\rho_k}{k!} \|h_k\|_k^2 < \infty,$$

where  $\|h_0\|_0 := |h_0|$  and

$$\|h_k\|_k := \left( \sum_{i_1=0}^L \dots \sum_{i_k=0}^L |h_k(i_1, \dots, i_k)|^2 \right)^{1/2}$$

for  $k = 1, 2, \dots$ . Then any function  $g : \mathcal{U} \rightarrow \mathbb{R}$  of the form

$$g(u_{0:L}) = h_0 + \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \in \{0, \dots, L\}} h_k(i_1, \dots, i_k) \prod_{j=1}^k u_{i_j}$$

is an element of  $\mathcal{H}_{\mathcal{U}}$  [23].

Removing the restriction  $A_1 = \dots = A_L = 0$  yields a broader class of nonlinear systems that subsumes the models studied in [23].

3) *State-space models*: Consider a Hammerstein state-space model of the form

$$\begin{aligned} x_{t+1} &= Ax_t + B\psi_1(u_t), \\ y_t &= Cx_t + D\psi_2(u_t), \end{aligned} \quad (10)$$

where  $\psi_1, \psi_2 : \mathcal{U} \rightarrow \mathbb{R}^q$  are two (nonlinear) functions. We assume that the state  $x_t$  takes values in  $\mathbb{R}^n$ , so  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{n \times q}$ ,  $C \in \mathbb{R}^{p \times n}$ , and  $D \in \mathbb{R}^{p \times q}$ . We can obtain an input-output representation (1) from the state-space representation (10) as follows.

Introduce the Markov parameters  $M_j := CA^jB \in \mathbb{R}^{p \times q}$  for  $j \geq 0$ . Define the  $L$ -step controllability matrix  $\mathcal{C}_L$ , the  $L$ -step observability matrix  $\mathcal{O}_L$ , the reversed  $L$ -step controllability matrix  $\Delta_L$ , and the modified block Toeplitz operator  $\tilde{\mathcal{T}}_L$  as

$$\begin{aligned} \mathcal{C}_L &:= [B \quad AB \quad A^2B \quad \dots \quad A^{L-1}B], \\ \mathcal{O}_L &:= \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{bmatrix}, \\ \Delta_L &:= [A^{L-1}B \quad \dots \quad AB \quad B], \\ \tilde{\mathcal{T}}_L &= \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ M_0 & 0 & \dots & 0 & 0 \\ M_1 & M_0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ M_{L-2} & M_{L-2} & \dots & M_0 & 0 \end{bmatrix}. \end{aligned} \quad (11)$$

Let us assume that  $L \geq n$  is such that  $\text{rank}(\mathcal{O}_L) = n$ . Then from (10), for  $k = 0, \dots, L$ , we have

$$\begin{aligned} y_{t-L+k} &= CA^k x_{t-L} + \sum_{j=0}^{k-1} M_j \psi_1(u_{t-L+k-1-j}) \\ &\quad + D\psi_2(u_{t-L+k}). \end{aligned} \quad (12)$$

Define the vectors

$$\begin{aligned} Y_t &:= \begin{bmatrix} y_{t-L} \\ \vdots \\ y_{t-1} \end{bmatrix} \in \mathbb{R}^{pL}, \\ E_t &:= \begin{bmatrix} \psi_1(u_{t-L}) \\ \vdots \\ \psi_1(u_{t-1}) \end{bmatrix} \in \mathbb{R}^{qL}, \\ F_t &:= \begin{bmatrix} \psi_2(u_{t-L}) \\ \vdots \\ \psi_2(u_{t-1}) \end{bmatrix} \in \mathbb{R}^{qL}. \end{aligned}$$

Then, stacking the equations in (12) for  $k = 0$  to  $k = L - 1$ , we have

$$Y_t = \mathcal{O}_L x_{t-L} + \tilde{\mathcal{T}}_L E_t + (I_L \otimes D) F_t.$$

Since  $\text{rank}(\mathcal{O}_L) = n$ , we can solve for  $x_{t-L}$ :

$$x_{t-L} = \mathcal{O}_L^\dagger \left( Y_t - \tilde{\mathcal{T}}_L E_t - (I_L \otimes D) F_t \right). \quad (13)$$

On the other hand, when  $k = L$ , we have

$$y_t = CA^L x_{t-L} + \sum_{j=0}^{L-1} M_j \psi_1(u_{t-1-j}) + D\psi_2(u_t).$$

Plugging in the expression for  $x_{t-L}$  from (13),

$$\begin{aligned} y_t &= CA^L \mathcal{O}_L^\dagger \left( Y_t - \tilde{\mathcal{T}}_L E_t - (I_L \otimes D) F_t \right) \\ &\quad + \sum_{j=0}^{L-1} M_j \psi_1(u_{t-1-j}) + D\psi_2(u_t). \end{aligned}$$

The matrix  $Q := CA^L \mathcal{O}_L^\dagger \in \mathbb{R}^{p \times pL}$  can be written in block form as  $Q = [Q_{L-1}, \dots, Q_0]$  with  $Q_i \in \mathbb{R}^{p \times p}$  for  $i = 0, \dots, L-1$ . Setting  $A_k := -Q_k$  for  $k = 1, \dots, L-1$  and moving them to the LHS of the above equation, we have

$$\begin{aligned} y_t + \sum_{k=1}^L A_k y_{t-k} &= \sum_{j=0}^{L-1} M_j \psi_1(u_{t-1-j}) - Q \tilde{\mathcal{T}}_L E_t \\ &\quad - Q(I_L \otimes D) F_t + D\psi_2(u_t). \end{aligned}$$

With  $M = [M_{L-1}, \dots, M_0] \in \mathbb{R}^{p \times mL}$ , we can write the RHS as a function of on  $u_{t-L:t}$  as

$$\begin{aligned} y_t + \sum_{k=1}^L A_k y_{t-k} &= \underbrace{\begin{bmatrix} M - Q \tilde{\mathcal{T}}_L & -Q(I_L \otimes D) & D \end{bmatrix}}_{=:S} \underbrace{\begin{bmatrix} E_t \\ F_t \\ \psi_2(u_t) \end{bmatrix}}_{=: \psi(u_{t-L:t})} \\ &=: g(u_{t-L:t}), \end{aligned}$$

where  $S \in \mathbb{R}^{p \times q(2L+1)}$ ,  $\psi(u_{t-L:t}) \in \mathbb{R}^{q(2L+1)}$ , and  $g$  defined above is a mapping from  $\mathcal{U}$  into  $\mathcal{Y} = \mathbb{R}^p$ . Assume now that  $\psi_1, \psi_2$  are elements of some  $\mathbb{R}^q$ -valued RKHS  $\mathcal{H}$  on  $\mathcal{U}$ . Thus, by definition, the evaluation maps  $\psi_1 \mapsto \psi_1(u)$  and  $\psi_2 \mapsto \psi_2(u)$  are bounded for each  $u \in \mathcal{U}$ . Consider the direct-sum Hilbert space  $\mathcal{K} := \mathcal{H}^{\oplus(2L+1)}$  with inner product  $\langle h, h' \rangle_{\mathcal{H}^{\oplus(2L+1)}} = \sum_{j=0}^{2L} \langle h_j, h'_j \rangle_{\mathcal{H}}$  for  $h = h_0 \oplus \dots \oplus h_{2L}$ . The function  $g$  defined above is an element of the linear space  $\mathcal{F}$  of functions  $f : \mathcal{U} \rightarrow \mathcal{Y}$  of the form

$$f(u_{0:L}) = S_0 h_0(u_0) + \sum_{j=1}^L S_j h_j(u_j) + \sum_{j=L+1}^{2L} S_j h_j(u_{j-L}),$$

where  $h = h_0 \oplus \dots \oplus h_{2L}$  ranges over  $\mathcal{K}$  and where  $S_0, S_1, \dots, S_{2L}$  are linear operators from  $\mathbb{R}^q$  into  $\mathcal{Y} = \mathbb{R}^p$ . For each  $u_{0:L}$ , the evaluation map  $f \mapsto f(u_{0:L})$  is continuous on  $\mathcal{F}$ . Thus,  $g$  belongs to some  $\mathcal{Y}$ -valued RKHS on  $\mathcal{U}$ .

#### IV. BEHAVIOR REPRESENTER THEOREM VIA MINIMUM NORM INTERPOLATION

Let  $(u_t^d, y_t^d)_{t=0}^{T+L-1}$  be a finite input-output trajectory generated by an unknown system of the form (1). In other words, there exists an unknown  $(L+1)$ -tuple  $(A_1, \dots, A_L, g)$ , such that

$$\begin{bmatrix} u_{0:T+L-1}^d \\ y_{0:T+L-1}^d \end{bmatrix} \in \mathcal{B}(A_1, \dots, A_L, g)|_{T+L-1}.$$

The problem of data-driven behavioral modeling is to reconstruct the unknown length- $L$  behavior segment  $\mathcal{B}(A_1, \dots, A_L, g)|_L$  from the data without explicitly estimating  $A_1, \dots, A_L, g$ .

Let  $f$  be the regression representation of  $(A_1, \dots, A_L, g)$  as in (3). Using the definitions of  $z_t$  in (2) and  $y_{t+} := y_{t+L}$ , we can represent the data  $(u_t^d, y_t^d)_{t=0}^{T+L-1}$  equivalently by  $(z_t^d, y_{t+}^d)_{t=0}^{T-1}$ , such that  $f(z_t^d) = y_{t+}$  holds for all  $t = 0, \dots, T-1$ . In other words, for each  $t = 0, \dots, T-1$ ,  $(\kappa_{\mathcal{Z}}(\cdot, z_t^d)^*, y_{t+}^d)$  is a valid i/o pair for (3). In view of the equivalence (8), the question we would like to answer is whether we can reconstruct the set of *all* valid i/o pairs of the unknown  $f$  from the given data.

The following result is easy to establish:

**Theorem 3.** *For any collection of real coefficients  $c_0, \dots, c_{T-1}$ ,  $(\sum_{j=0}^{T-1} c_j \kappa_{\mathcal{Z}}(\cdot, z_j^d)^*, \sum_{j=0}^{T-1} c_j y_{j+}^d)$  is a valid i/o pair for the model (3).*

Theorem 3 shows that any linear combination of  $(\kappa_{\mathcal{Z}}(\cdot, z_t^d)^*, y_t^d)_{t=0}^{T-1}$  is a valid i/o pair of (3). We now analyze the reverse direction by relating it to minimum-norm interpolation in a vector-valued RKHS [24]. Let finite input-output data  $\{(z_i, y_i)\}_{i=0}^{N-1} \subset \mathcal{Z} \times \mathcal{Y}$  be given and consider the following problem:

$$\begin{aligned} & \text{minimize } \|f\|_{\mathcal{H}_{\mathcal{Z}}}^2 \\ & \text{subject to } f(z_j) = y_j, \quad j = 0, \dots, N-1. \end{aligned} \quad (14)$$

Define the *sampling operator*  $S_N : \mathcal{H}_{\mathcal{Z}} \rightarrow \mathcal{Y}^N$  by  $S_N f = (f(z_0), \dots, f(z_{N-1}))$ . By [24, Theorem 3], if  $(y_0, \dots, y_{N-1}) \in \text{range}(S_N)$ , then the minimum norm interpolation problem in (14) has a unique solution given by

$$f_N = \sum_{j=0}^{N-1} \kappa_{\mathcal{Z}}(\cdot, z_j) v_j, \quad (15)$$

where  $(v_0, \dots, v_{N-1}) \in \mathcal{Y}^N$  solves the system of equations

$$\sum_{l=0}^{N-1} \kappa_{\mathcal{Z}}(z_j, z_l) v_l = y_j, \quad j = 0, \dots, N-1. \quad (16)$$

We can express this more succinctly as follows.

Since  $\mathcal{Y} = \mathbb{R}^p$ , for each  $z \in \mathcal{Z}$  we can view  $\kappa_{\mathcal{Z}}(\cdot, z)$  as a mapping from  $\mathbb{R}^p$  into  $\mathcal{H}_{\mathcal{Z}}$ , i.e., for each  $v \in \mathbb{R}^p$ ,  $\kappa_{\mathcal{Z}}(\cdot, z)v \in \mathcal{H}_{\mathcal{Z}}$ . Define the mapping  $\Phi_N : \mathbb{R}^{pN} \rightarrow \mathcal{H}_{\mathcal{Z}}$  as

$$\Phi_N \bar{v} = \sum_{j=0}^{N-1} \kappa_{\mathcal{Z}}(\cdot, z_j) v_j,$$

where  $\bar{v} = (v_0, \dots, v_{N-1}) \in \mathbb{R}^{pN}$ . In particular,  $f_N = \Phi_N \bar{v}$ , where  $\bar{v}$  is the solution of (16). Next, define the block kernel matrix  $K_N := \Phi_N^* \Phi_N \in \mathbb{R}^{pN \times pN}$  whose blocks are given by  $[K_N]_{ij} = \kappa_{\mathcal{Z}}(z_i, z_j)$  for  $i, j = 0, \dots, N-1$ , as well as the block row vector  $k_N(z) := \kappa_{\mathcal{Z}}(\cdot, z)^* \Phi_N \in \mathbb{R}^{p \times pN}$  with blocks  $[k_N(z)]_j = \kappa_{\mathcal{Z}}(z, z_j)$  for  $j = 0, \dots, N-1$ . Using the reproducing property and the above definitions, we can write

$$\begin{aligned} f_N(z) &= \kappa_{\mathcal{Z}}(\cdot, z)^* f_N \\ &= \kappa_{\mathcal{Z}}(\cdot, z)^* \Phi_N K_N^\dagger \bar{v} \end{aligned}$$

$$= k_N(z) K_N^\dagger Y_N, \quad (17)$$

where  $Y_N := [y_0^T, \dots, y_{N-1}^T]^T$ . Finally, for each  $z$  define  $\Sigma_N(z) \in \mathbb{R}^{p \times p}$  as

$$\Sigma_N(z) := \kappa_{\mathcal{Z}}(z, z) - k_N(z) K_N^\dagger k_N(z)^*. \quad (18)$$

We now present two key lemmas. The first one is a structural characterization of  $\Sigma_N$ :

**Lemma 4.** *For each  $z$ , the operator  $\Sigma_N(z)$  is symmetric and positive semi-definite, and  $\Sigma_N(z) = 0$  iff  $\text{range}(\kappa_{\mathcal{Z}}(\cdot, z)) \subseteq \text{range}(\Phi_N)$ .*

The second one is an extension of a result of Liang and Recht [25, Lemma 1] to the minimum-norm interpolation problem in the vector-valued RKHS  $\mathcal{H}_{\mathcal{Z}}$ :

**Lemma 5.** *Suppose that the problem (14) admits unique solutions  $f_N$  and  $f_{N+1}$  given the respective data  $\{(z_i, y_i)\}_{i=0}^{N-1}$  and  $\{(z_i, y_i)\}_{i=0}^{N-1} \cup \{(z_N, y_N)\}$ . Then the following holds:*

- 1) If  $\Sigma_N(z_N) = 0$ , then  $y_N - f_N(z_N) = 0$ .
- 2) If  $\Sigma_N(z_N) \succ 0$ , we have

$$\begin{aligned} & \|f_{N+1}\|_{\mathcal{H}_{\mathcal{Z}}}^2 - \|f_N\|_{\mathcal{H}_{\mathcal{Z}}}^2 \\ &= \left\| \Sigma_N(z_N)^{-1/2} (y_N - f_N(z_N)) \right\|_{\mathcal{Y}}^2. \end{aligned} \quad (19)$$

Lemma 5 characterizes the error incurred when we use  $\hat{y}_N = f_N(z_N)$  as a *predictor* of  $y_N = f_{N+1}(z_N)$ , where  $f_N$  is the solution to the minimum norm interpolation problem on  $\{(z_i, y_i)\}_{i=0}^{N-1}$ . In particular, the prediction is exact when  $\Sigma_N(z_N) = 0$ . If  $\Sigma_N(z_N)$  is nonzero but still positive definite, the identity (19) expresses the squared norm of the weighted prediction error  $\Sigma_N(z_N)^{-1/2} (y_N - \hat{y}_N)$  in terms of the norms of  $f_{N+1}$  and  $f_N$ .

In a scalar-valued RKHS,  $\Sigma_N(z)$  is equal to

$$\begin{aligned} & \text{dist}^2(\kappa_{\mathcal{Z}}(\cdot, z), \text{span}(\kappa_{\mathcal{Z}}(\cdot, z_0), \dots, \kappa_{\mathcal{Z}}(\cdot, z_{N-1}))) \\ &= \min_{h \in \text{span}(\kappa_{\mathcal{Z}}(\cdot, z_0), \dots, \kappa_{\mathcal{Z}}(\cdot, z_{N-1}))} \|\kappa_{\mathcal{Z}}(\cdot, z) - h\|_{\mathcal{H}_{\mathcal{Z}}}^2, \end{aligned} \quad (20)$$

the squared distance from  $\kappa_{\mathcal{Z}}(\cdot, z)$  to the linear span of kernel functions centered at the observed samples [25]. While Lemma 5 applies to *any* vector-valued RKHS of  $\mathcal{Y}$ -valued functions on  $\mathcal{H}_{\mathcal{Z}}$ , a natural choice of the operator-valued kernel could be  $\kappa_{\mathcal{Z}}(z_1, z_2) = \kappa_{\mathcal{Z}}^s(z_1, z_2) I_{\mathcal{Y}}$ , where  $\kappa_{\mathcal{Z}}^s$  is a scalar-valued kernel function. Let  $K_N^s = [\kappa_{\mathcal{Z}}^s(z_i, z_j)]_{i,j=0}^{N-1} \in \mathbb{R}^{N \times N}$  be the Gram matrix defined by the scalar-valued kernel  $\kappa_{\mathcal{Z}}^s$  on the points  $z_0, \dots, z_{N-1}$ , let  $k_N^s(z) \in \mathbb{R}^{1 \times N}$  be the row vector with  $[k_N^s(z)]_j = \kappa_{\mathcal{Z}}^s(z, z_j)$ ,  $K_N = K_N^s \otimes I_{\mathcal{Y}}$ , and  $K_N(z_N) = k_N^s(z_N) \otimes I_{\mathcal{Y}}$ . Finally, define

$s_N := \text{dist}\{\text{span}(\kappa_{\mathcal{Z}}^s(\cdot, z_0), \dots, \kappa_{\mathcal{Z}}^s(\cdot, z_{N-1})), \kappa_{\mathcal{Z}}^s(\cdot, z_N)\}$ , where the distance is computed in the scalar-valued RKHS induced by  $\kappa_{\mathcal{Z}}^s$ , cf. (20). Then, using the properties of the Kronecker product  $\otimes$ , we can compute  $\Sigma_N(z_N)$  as follows:

$$\begin{aligned} & \Sigma_N(z_N) \\ &= \kappa_{\mathcal{Z}}(z_N, z_N) - K_N(z_N) K_N^\dagger K_N(z_N)^* \\ &= \kappa_{\mathcal{Z}}^s(z_N, z_N) I_{\mathcal{Y}} \\ &\quad - (k_{N-1}^s(z_N) \otimes I_{\mathcal{Y}})^* (K_{N-1}^s \otimes I_{\mathcal{Y}})^\dagger (k_{N-1}^s(z_N) \otimes I_{\mathcal{Y}}) \end{aligned}$$

$$\begin{aligned}
&= (\kappa_{\mathcal{Z}}^s(z_N, z_N) - k_{N-1}^s(z_N)^T (K_{N-1}^s)^\dagger k_{N-1}^s(z_N)) I_{\mathcal{Y}} \\
&= s_N^2 I_{\mathcal{Y}}.
\end{aligned}$$

Hence, the equality in Lemma 5 reduces to

$$s_N^2 \left( \|f_{N+1}\|_{\mathcal{H}_{\mathcal{Z}}}^2 - \|f_N\|_{\mathcal{H}_{\mathcal{Z}}}^2 \right) = \|y_N - f_N(z_N)\|_{\mathcal{Y}}^2.$$

This indicates that, given a new  $(z_N, y_N)$ , if  $s_N = 0$ , we have  $\|y_N - f_N(z_N)\|_{\mathcal{Y}}^2 = 0$ . Given Lemmas 4 and 5, the following is immediate:

**Theorem 6** (Behavior Representer Theorem). *Let  $(z_t^d, y_{t+}^d)_{t=0}^{T-1}$  be a length- $T$  trajectory of regression and output vectors for an unknown  $f_* \in \mathcal{H}_{\mathcal{Z}}$ , i.e.,*

$$y_{t+}^d = f_*(z_t^d) \quad \text{for each } t = 0, \dots, T-1.$$

Let  $[u_{0:L}^r, y_{0:L}^r]^r$  be an element of  $\mathcal{B}(f_*)|_L$ , and let  $(z_0, y_{0+})$  be computed according to (7). Then the following holds:

- 1) If  $\Sigma_T(z_0) = 0$ , then  $y_{0+} = k_T(z_0)K_T^\dagger Y_T$ , where  $K_T$  and  $Y_T$  are computed from the data according to (17) and  $Y_T := [(y_{0+}^d)^r, \dots, (y_{(T-1)+}^d)^r]^r$ .
- 2) If  $\Sigma_T(z_0) \succ 0$ , then

$$\begin{aligned}
&\|\Sigma_T(z_0)^{-1/2}(y_{0+} - k_T(z_0)K_T^\dagger Y_T)\|_{\mathcal{Y}}^2 \\
&= \|f_{T+1}\|_{\mathcal{H}_{\mathcal{Z}}}^2 - \|f_T\|_{\mathcal{H}_{\mathcal{Z}}}^2 \\
&\leq \|f_*\|_{\mathcal{H}_{\mathcal{Z}}}^2 - \|f_T\|_{\mathcal{H}_{\mathcal{Z}}}^2,
\end{aligned} \tag{21}$$

where  $f_T$  (respectively,  $f_{T+1}$ ) is the minimum-norm interpolator of  $\{(z_t^d, y_{t+}^d)\}_{t=0}^{T-1}$  (respectively, of  $\{(z_t^d, y_{t+}^d)\}_{t=0}^{T-1} \cup \{(z_0, y_{0+})\}$ ).

The condition  $\Sigma_T(z_0) = 0$  describes the setting when exact reconstruction is possible. By Lemma 4, this will be the case for all pairs  $(z, y_+ = f_*(z))$  for which  $\text{range}(\kappa_{\mathcal{Z}}(\cdot, z)) \subseteq \text{range}(\Phi_T)$ . If  $z_0$  does not satisfy this condition but  $\Sigma_T(z_0) \succ 0$ , we can quantify the reconstruction error using (21). For LTI systems, a sufficient condition for the existence of unique minimum-norm interpolating solutions is

$$\sum_{t=0}^{T-1} z_t^d (z_t^d)^r \succ 0,$$

which is the classical persistence of excitation condition [21, 22]. Exact reconstruction is possible when  $z_0 \in \text{span}(z_0^d, \dots, z_{T-1}^d)$ . In the nonlinear setting, the condition  $\Sigma_T(z_0) = 0$  plays an analogous role via Lemma 4.

It is useful to compare Theorem 6 with the fundamental lemma for LTI systems (cf. Section II). The latter characterizes behaviors through the image of the Hankel matrix built from observed input-output trajectories. Theorem 6 of this section is conceptually analogous, but is phrased in terms of a different data representation based on regression vectors, which are formed from inputs and past outputs. A related perspective on the fundamental lemma via regression in reproducing kernel Hilbert spaces was recently explored in [15].

## V. SUBSPACE IDENTIFICATION IN THE VECTOR-VALUED RKHS SETTING

We now revisit systems that arise from state-space representations, as in Section III-C3. We will focus on a particular class of such systems, namely ones that can be represented as

$$\begin{aligned}
x_{t+1} &= Ax_t + B\phi(u_t), \\
y_t &= Cx_t + D\phi(u_t)
\end{aligned} \tag{22}$$

where  $\phi: \mathcal{U} \rightarrow \mathbb{R}^q$  is a mapping from the input space  $\mathcal{U} = \mathbb{R}^m$  into  $\mathbb{R}^q$ .

Given  $(A, B, C, D)$ , we construct the  $L$ -step controllability matrix  $\mathcal{C}_L$ , the  $L$ -step observability matrix  $\mathcal{O}_L$ , the reversed  $L$ -step controllability matrix  $\Delta_L$ , and the  $L$ -step modified Toeplitz matrix  $\tilde{\mathcal{T}}_L$  exactly as in (11). The  $L$ -step Toeplitz matrix is given by

$$\mathcal{T}_L := \tilde{\mathcal{T}}_L + I_L \otimes D.$$

Let  $\{(u_t^d, y_t^d)\}_{t=0}^{T-1}$  denote input/output data of length  $T$  collected from measurements of (22) starting from some initial condition  $x_0 \in \mathbb{R}^n$ . In the LTI case (i.e., when  $q = m$  and  $\phi$  is the identity map), subspace identification methods [17] allow one to reconstruct, under certain regularity conditions, the state trajectory  $x_L, \dots, x_{T-L}$  and the observability matrix  $\mathcal{O}_L$  directly from the input/output data without knowledge of the system matrices  $A, B, C, D$ .

In this section, we show that these methods can be extended to the set-up of (22) when  $\phi$  is an element of a suitable vector-valued RKHS on  $\mathcal{U}$ . Moreover, we obtain a result in the spirit of the fundamental lemma for LTI systems, namely that the set of all valid length- $L$  input/output trajectories of (22) can be reconstructed directly from the data  $\{(u_t^d, y_t^d)\}_{t=0}^{T-1}$  without an intermediate system identification step.

### A. The construction of a vector-valued RKHS

There are various ways of instantiating the state-space model (22) in the vector-valued RKHS framework presented in Section III-C, i.e., choosing a suitable  $\mathbb{R}^q$ -valued RKHS  $\mathcal{H}_\kappa$  on the input space  $\mathcal{U}$  with reproducing kernel  $\kappa$  so that  $\phi \in \mathcal{H}_\kappa$ . Perhaps the simplest one is to define the operator-valued map  $\kappa: \mathcal{U} \times \mathcal{U} \rightarrow \mathcal{L}(\mathbb{R}^q)$  by

$$\kappa(u, u') := \phi(u) \otimes \phi(u'),$$

which is readily seen to be of the positive type since, for any  $n$ , any  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ , any  $u_1, \dots, u_n \in \mathcal{U}$ , and any  $v \in \mathbb{R}^q$ ,

$$\begin{aligned}
&\sum_{i,j=1}^n \alpha_i \alpha_j \langle \kappa(u_i, u_j) v, v \rangle_{\mathbb{R}^q} \\
&= \sum_{i,j=1}^n \alpha_i \alpha_j \langle \phi(u_i), v \rangle_{\mathbb{R}^q} \langle \phi(u_j), v \rangle_{\mathbb{R}^q} \\
&= \left( \sum_{i=1}^n \alpha_i \langle \phi(u_i), v \rangle_{\mathbb{R}^q} \right)^2 \\
&\geq 0.
\end{aligned}$$

By [20, Prop. 2.3], there is a unique  $\mathbb{R}^q$ -valued RKHS  $\mathcal{H}_\kappa$  of functions on  $\mathcal{U}$  with reproducing kernel  $\kappa$ . The kernel  $\kappa$  has the convenient property

$$\langle \phi(u), \phi(u') \rangle_{\mathbb{R}^q} = \text{Tr}(\kappa(u, u')). \quad (23)$$

### B. Subspace identification using RKHS methods

We now have all the ingredients in place for putting together a subspace identification framework analogous to the one for LTI systems [17].

Let the input/output data  $\{(u_t^d, y_t^d)\}_{t=0}^{T-1}$  be given. Introduce the Hankel-type block matrices

$$\begin{bmatrix} Y_p \\ Y_f \end{bmatrix} := \begin{bmatrix} y_0^d & \cdots & y_{T-2L}^d \\ \vdots & & \vdots \\ y_{L-1}^d & \cdots & y_{T-L-1}^d \\ y_L^d & \cdots & y_{T-L}^d \\ \vdots & & \vdots \\ y_{2L-1}^d & \cdots & y_{T-1}^d \end{bmatrix}. \quad (24)$$

and

$$\begin{bmatrix} U_p^\phi \\ U_f^\phi \end{bmatrix} := \begin{bmatrix} \phi(u_0^d) & \cdots & \phi(u_{T-2L}^d) \\ \vdots & & \vdots \\ \phi(u_{L-1}^d) & \cdots & \phi(u_{T-L-1}^d) \\ \phi(u_L^d) & \cdots & \phi(u_{T-L}^d) \\ \vdots & & \vdots \\ \phi(u_{2L-1}^d) & \cdots & \phi(u_{T-1}^d) \end{bmatrix}, \quad (25)$$

where p and f designates a partition into ‘‘past’’ and ‘‘future’’ data. Let  $H_p$  and  $H_f$  denote, respectively, the concatenations of  $U_p$  and  $Y_p$  and of  $U_f$  and  $Y_f$ :

$$H_p := \begin{bmatrix} U_p \\ Y_p \end{bmatrix}, \quad H_f := \begin{bmatrix} U_f \\ Y_f \end{bmatrix}.$$

Similar to the linear case, let  $\Pi$  denote the oblique projection of the rowspace of  $Y_f$  onto the rowspace of  $H_p$  along the rowspace of  $U_f$ :

$$\Pi := Y_f / \begin{matrix} H_p \\ U_f^\phi \end{matrix},$$

which satisfies  $\text{rank}(\Pi) \leq T - 2L + 1$ . Let the SVD of  $\Pi$  be given by

$$\Pi = [U_1 \quad U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \Sigma_1 V_1^T.$$

Following [17], let  $X_f := [x_L \quad x_{L+1} \quad \cdots \quad x_{T-L}]$ , where  $x_t$  is the state trajectory of (22) determined by the initial state  $x_0$  and the inputs  $u_t^d$ . In the LTI case,  $X_f$  can be recovered, up to a similarity transformation, via the formula  $X_f = \Sigma_1^{1/2} V_1^T$  (see, e.g., [17, Theorem 2, Chapter 2]). In fact, the same result holds in the present nonlinear case (the idea is to view  $v_k :=$

$\phi(u_k)$  as an input to an LTI system given by  $(A, B, C, D)$ ). Given the input/output data, let

$$U^\phi := \begin{bmatrix} \phi(u_0^d) & \cdots & \phi(u_{T-2L-n}^d) \\ \vdots & & \vdots \\ \phi(u_{2L-1}^d) & \cdots & \phi(u_{T-n}^d) \\ \phi(u_{2L}^d) & \cdots & \phi(u_{T-n}^d) \\ \vdots & & \vdots \\ \phi(u_{2L-1+n}^d) & \cdots & \phi(u_{T-1}^d) \end{bmatrix}, \quad (26)$$

and form the input Gram matrix of depth  $2L + n$  as  $K^u := (U^\phi)^T U^\phi$  with entries  $[K^u]_{ij} = \sum_{k=0}^{2L+n-1} \kappa(u_{i+k}^d, u_{j+k}^d)$ . The following theorem is a straightforward adaptation of subspace identification results for LTI systems:

**Theorem 7.** *Let  $\{u_i^d, y_i^d\}_{i=0}^{T-1}$  be a sequence of length  $T$  generated by the system (22), where  $(A, B)$  is controllable and  $(A, C)$  is observable. Suppose that the input sequence is such that  $\text{rank}(K^u) = (2L + n)q$ . Then,  $\Pi = \mathcal{O}_L X_f$  and  $X_f = \Sigma_1^{1/2} V_1^T$ .*

The process of computing the oblique projection and constructing the state vector can be carried out using Gram matrices computed from pairwise kernel evaluations  $\kappa$ , without explicitly using  $\phi$ . Specifically, define the Gram matrices  $K_p^u = (U_p^\phi)^T U_p^\phi$ ,  $K_f^u = (U_f^\phi)^T U_f^\phi$ , and  $K_p^y = Y_p^T Y_p$ , whose entries are

$$\begin{aligned} [K_p^u]_{ij} &= \sum_{k=0}^{L-1} \langle \phi(u_{i+k}^d), \phi(u_{j+k}^d) \rangle_{\mathbb{R}^M} \\ &\stackrel{(a)}{=} \sum_{k=0}^{L-1} \text{Tr}(\kappa(u_{i+k}^d, u_{j+k}^d)), \\ [K_f^u]_{ij} &= \sum_{k=0}^{L-1} \langle \phi(u_{i+L+k}^d), \phi(u_{j+L+k}^d) \rangle_{\mathbb{R}^M} \\ &\stackrel{(b)}{=} \sum_{k=0}^{L-1} \text{Tr}(\kappa(u_{i+L+k}^d, u_{j+L+k}^d)), \\ [K_p^y]_{ij} &= \sum_{k=0}^{L-1} y_{i+k}^d \text{ }^T y_{j+k}^d, \end{aligned}$$

respectively, where (a) and (b) follow from (23).

The oblique projection can be computed as

$$\begin{aligned} \Pi &= Y_f / \begin{matrix} H_p \\ U_f^\phi \end{matrix} \\ &= Y_f \begin{bmatrix} H_p^T & (U_f^\phi)^T \end{bmatrix} \left( \begin{bmatrix} H_p \\ U_f^\phi \end{bmatrix} \begin{bmatrix} H_p^T & (U_f^\phi)^T \end{bmatrix} \right)^\dagger \begin{bmatrix} H_p \\ 0 \end{bmatrix} \\ &= Y_f \left( \begin{bmatrix} H_p^T & (U_f^\phi)^T \end{bmatrix} \begin{bmatrix} H_p \\ U_f^\phi \end{bmatrix} \right)^\dagger \begin{bmatrix} H_p^T & (U_f^\phi)^T \end{bmatrix} \begin{bmatrix} H_p \\ 0 \end{bmatrix} \\ &= Y_f (K_p^u + K_p^y + K_f^u)^\dagger (K_p^u + K_p^y), \end{aligned} \quad (27)$$

where the second-to-last line follows from the identity  $x(x^T x)^\dagger = (x x^T)^\dagger x$ .

To construct the state vector, instead of directly performing SVD on  $\Pi$ , we construct  $U$ ,  $\Sigma$ , and  $V$  such that  $\Pi = U \Sigma V^T$

via eigendecomposition of  $\Pi^T\Pi$  and  $\Pi\Pi^T$ . Denote  $\overline{K}_{p,f} := K_p^u + K_p^y + K_f^u$  and  $\overline{K}_p := K_p^u + K_p^y$  and notice that we have

$$\begin{aligned}\Pi^T\Pi &= \overline{K}_p^T \left( \overline{K}_{p,f}^\dagger \right)^T (Y_f^T Y_f) \overline{K}_{p,f}^\dagger \overline{K}_p \\ &= \overline{K}_p^T \left( \overline{K}_{p,f}^\dagger \right)^T K_f^y \overline{K}_{p,f}^\dagger \overline{K}_p \\ &= \overline{K}_p^T \overline{K}_{p,f}^\dagger K_f^y \overline{K}_{p,f}^\dagger \overline{K}_p.\end{aligned}$$

Then, we can obtain  $(\sigma_i^{1/2}, v_i)$  as the eigenvalues and right eigenvectors of  $\Pi^T\Pi$ . Likewise, we have

$$\Pi\Pi^T = Y_f \overline{K}_{p,f}^\dagger \overline{K}_p \overline{K}_p^T \overline{K}_{p,f}^\dagger Y_f^T.$$

Denote  $\Gamma := \overline{K}_{p,f}^\dagger \overline{K}_p \overline{K}_p^T \overline{K}_{p,f}^\dagger$ . Then, we have

$$Y_f \Gamma Y_f^T (Y_f z_i) = \sigma_i^{1/2} (Y_f z_i) \Leftrightarrow \Gamma \underbrace{Y_f^T Y_f}_{=: K_f^y} \xi_i = \sigma_i^{1/2} \xi_i.$$

That is,  $\Pi\Pi^T = Y_f \Gamma Y_f^T$  has an eigenvector  $u_i = Y_f \xi_i$  associated with eigenvalue  $\sigma_i^{1/2}$  iff  $\xi_i$  is an eigenvector of  $\Gamma K_f^y$  corresponding to the same eigenvalue. With  $U = [u_1, \dots, u_n]$ ,  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ , and  $V = [v_1, \dots, v_n]$ , we can define the extended observability matrix  $\mathcal{O}_L$  and states  $X_f$  (up to a similarity transform) as

$$\mathcal{O}_L = U \Sigma^{1/2}, \quad X_f = V \Sigma^{1/2}.$$

The following corollary is immediate from the above process.

**Corollary 8** (Construction of state vectors). *Suppose that  $(A, B)$  is controllable and  $(A, C)$  is observable, i.e.,  $\text{rank}(\mathcal{O}_L) = \text{rank}(\mathcal{C}_L) = n$ . Let input-output data  $\{(u_t^d, y_t^d)\}_{t=0}^{T-1}$  of length  $T$  be given. Suppose the input sequence is such that  $\text{rank}(K_{2L+n}^u) = (2L+n)q$ . Let  $(\Sigma^{1/2}, V)$  be the eigenpair of  $\overline{K}_p^T (\overline{K}_{p,f}^{-1})^T K_f^y \overline{K}_{p,f}^{-1} \overline{K}_p$ , and  $(\Sigma^{1/2}, \Xi)$  be the eigenpair of  $\overline{K}_{p,f}^{-1} \overline{K}_p \overline{K}_p^T (\overline{K}_{p,f}^{-1})^T K_f^y$ . Define  $U = Y_f \Xi$ . Then the extended observability matrix is  $\mathcal{O}_L = U \Sigma^{1/2}$ , and the state sequence is  $X_f = V \Sigma^{1/2}$  (both up to a similarity transformation).*

The next result explicitly connects subspace identification to an RKHS version of the fundamental lemma for nonlinear systems admitting state-space realization (22).

**Theorem 9.** *Consider the state-space model (22), where  $(A, B)$  is controllable and  $(A, C)$  is observable. Let input-output data  $\{(u_t^d, y_t^d)\}_{t=0}^{T-1}$  of length  $T$  be given, and suppose the input Gram matrix is such that  $\text{rank}(K_{2L+n}^u) = (2L+n)q$  and that  $L$  is chosen so that  $\mathcal{O}_L$  has full column rank. Then a length- $2L$  sequence  $\begin{bmatrix} u_{0:2L-1} \\ y_{0:2L-1} \end{bmatrix}$  is a valid input/output trajectory of the system (22) if and only if there exists  $\xi \in \mathbb{R}^{T-2L+1}$  such that*

$$\begin{bmatrix} k_p^u \\ k_p^y \end{bmatrix} = \begin{bmatrix} K_p^u \\ K_p^y \end{bmatrix} \xi \quad \text{and} \quad \begin{bmatrix} k_f^u \\ k_f^y \end{bmatrix} = \begin{bmatrix} K_f^u \\ K_f^y \end{bmatrix} \xi,$$

where, for  $\Phi_{0:2L-1} := [\phi(u_0)^T, \dots, \phi(u_{2L-1})^T]^T$ , the vectors  $k_p^u = (U_p^\phi)^T \Phi_{0:2L-1}$  and  $k_f^u = (U_f^\phi)^T \Phi_{0:2L-1}$  have entries given by  $[k_p^u]_j = \sum_{r=0}^{L-1} \kappa(u_{j+r}^d, u_r)$ ,  $[k_f^u]_j = \sum_{r=0}^{L-1} \kappa(u_{j+L+r}^d, u_{L+r})$ ; and  $k_p^y = Y_p^T y_{0:L-1}$ ,  $k_f^y = Y_f^T y_{L:2L-1}$  are the Euclidean inner products of outputs.

In the context of the above theorem, we can view  $(K_p^u, K_p^y, K_f^u, K_f^y)$  as kernel matrices of the ‘‘offline’’ training data generated by (22). By splitting the kernel vector of the ‘‘online’’ testing sequence  $\Phi_{0:2L-1}$  into two parts (past and future), we test whether one can interpolate the kernel matrices of past and future using the same coefficients, where the initial length- $L$  segment of the testing sequence specifies the initial condition for the subsequent length- $L$  segment. A similar observation for LTI systems on the specification of initial condition has been given in [5, Proposition 1]. In the proof of Theorem 9, we further show that the trajectory passes through a state  $\hat{x}_L$  at time  $L$  that is a linear combination of states from the training data, i.e.,  $\hat{x}_L = X_f \xi$  for some  $\xi$ .

In light of Theorem 9, we can compute the predictor  $\hat{y}_{L:2L-1}$  via

$$\hat{y}_{L:2L-1} = (Y_f^T)^\dagger k_f^y = (Y_f^T)^\dagger K_f^y \xi,$$

where  $\xi$  is the solution of the following equation,

$$\begin{bmatrix} k_p^u \\ k_p^y \\ k_f^u \end{bmatrix} = \begin{bmatrix} K_p^u \\ K_p^y \\ K_f^u \end{bmatrix} \xi.$$

The interpolation-based method in Section IV characterizes the future output  $Y_f$  using kernelized offline data to make predictions regarding the online data. On the other hand, the realization-based method aims to capture the output  $y_{t+L-1}$  as linear functions of states  $X_f$  and kernelized input vectors  $\phi(u_t), \dots, \phi(u_{t+L-1})$ . The preference between the two approaches depends on the memory length  $L$  and the state dimension  $n$ . E.g., in the regime  $L \gg n$ , the states serve as an efficient representation of the system’s history.

## VI. CONCLUSION

In this paper, we have put forward a behavioral framework for modeling a class of nonlinear systems in a vector-valued RKHS. This formulation is rich enough to cover LTI systems as well as nonlinear systems modeled by Volterra series, autoregressive models based on Volterra series, and Hammerstein-type state-space models. Using this framework, we have analyzed two methods for data-driven modeling of such systems, minimum-norm interpolation and subspace identification. We have clarified the role of various structural assumptions on the system, the data (both offline and online), and the vector-valued RKHS that represents the nonlinear aspects of the system. More broadly, this work expands the scope of behavioral systems theory to nonlinear systems. In doing so, it reinforces the conceptual shift underlying data-driven control: rather than aiming to recover the state evolution, one can directly actualize the desired behaviors using observed trajectories.

## APPENDIX A PROOF OF LEMMA 2

Let  $\|\cdot\|_{\mathcal{V}}$  and  $\|\cdot\|_{\mathcal{W}}$  be the Hilbert-space norms on  $\mathcal{V}$  and  $\mathcal{W}$  induced by their respective inner products. Consider the

evaluation functional  $\delta_v : \mathcal{L}(\mathcal{V}, \mathcal{W}) \rightarrow \mathcal{W}$  given by  $\delta_v(A) = Av$ . It is bounded since

$$\begin{aligned} \|\delta_v(A)\|_{\mathcal{W}} &= \|Av\|_{\mathcal{W}} \\ &\leq \|A\|_{\mathcal{L}(\mathcal{V}, \mathcal{W})} \|v\|_{\mathcal{V}} \\ &\leq \|A\|_{\mathcal{H}} \|v\|_{\mathcal{V}}, \end{aligned}$$

where the last step follows from the relation

$$\begin{aligned} \|A\|_{\mathcal{L}(\mathcal{V}, \mathcal{W})} &= \sup_{v \in \mathcal{V}, \|v\|_{\mathcal{V}}=1} \sqrt{\langle Av, Av \rangle_{\mathcal{W}}} \\ &= \sup_{v \in \mathcal{V}, \|v\|_{\mathcal{V}}=1} \sqrt{\langle v, A^*Av \rangle_{\mathcal{V}}} \\ &\leq \|A\|_{\mathcal{H}}. \end{aligned}$$

Hence,  $\mathcal{H}$  is an RKHS. We next show that  $\kappa(v, v') = \langle v, v' \rangle_{\mathcal{V}} I_{\mathcal{W}}$  is the reproducing kernel. It is obvious that  $\kappa$  is of the positive type. To show it satisfies the reproducing property, notice that for any  $v \in \mathcal{V}$  and  $w \in \mathcal{W}$  we have

$$\begin{aligned} \langle Av, w \rangle_{\mathcal{W}} &= \text{Tr}(A^*(w \otimes v)) \\ &= \langle A, w \otimes v \rangle_{\mathcal{H}}, \end{aligned}$$

where  $w \otimes v \in \mathcal{H}$  is the rank-one linear operator  $(w \otimes v)v' = \langle v, v' \rangle_{\mathcal{V}} w$ .

On the other hand, for  $w \in \mathcal{W}$ , we have

$$\kappa(v', v)w = \langle v', v \rangle_{\mathcal{V}} w = (w \otimes v')v.$$

Hence, we have

$$\langle A, \kappa(\cdot, v)w \rangle_{\mathcal{H}} = \langle Av, w \rangle_{\mathcal{W}}.$$

That is,  $\kappa$  satisfies the reproducing property.

## APPENDIX B PROOFS FOR RESULTS IN SECTION IV

### A. Proof of Theorem 3

Let  $c_0, \dots, c_{T-1} \in \mathbb{R}$  be given. For each  $j = 0, \dots, T-1$ , the relation

$$\langle y_{j+}^d, v \rangle_{\mathcal{Y}} = \langle f, \kappa_{\mathcal{Z}}(\cdot, z_j^d)v \rangle_{\mathcal{H}_{\mathcal{Z}}}$$

holds for all  $v \in \mathcal{Y}$ . Multiplying both sides by  $c_j$  and summing over  $j = 0, \dots, T-1$ , we have

$$\begin{aligned} \left\langle \sum_{j=0}^{T-1} c_j y_{j+}^d, v \right\rangle_{\mathcal{Y}} &= \left\langle f, \sum_{j=0}^{T-1} c_j \kappa_{\mathcal{Z}}(\cdot, z_j^d)v \right\rangle_{\mathcal{H}_{\mathcal{Z}}} \\ &= \left\langle \sum_{j=0}^{T-1} c_j f(z_j^d), v \right\rangle_{\mathcal{H}_{\mathcal{Z}}}, \end{aligned}$$

for all  $v \in \mathcal{Y}$ , where the last line follows from the reproducing property. Hence, the pair  $\left( \sum_{j=0}^{T-1} c_j \kappa_{\mathcal{Z}}(\cdot, z_j^d), \sum_{j=0}^{T-1} c_j y_{j+}^d \right)$  satisfies Eq. (3) as claimed.

### B. Proof of Lemma 4

Define the following subspace  $\mathcal{H}_N$  of  $\mathcal{H}_{\mathcal{Z}}$ :

$$\begin{aligned} \mathcal{H}_N &:= \text{span} \{ \kappa_{\mathcal{Z}}(\cdot, z_i)v, i = 0, \dots, N-1, v \in \mathcal{Y} \} \\ &= \text{range}(\Phi_N). \end{aligned} \quad (28)$$

Let  $\Pi_N$  denote the orthogonal projection onto  $\mathcal{H}_N$  and notice that  $\Pi_N = \Phi_N (\Phi_N^* \Phi_N)^{\dagger} \Phi_N^*$ . Using this in the definition of  $\Sigma_N(z)$  in (18), we get

$$\begin{aligned} \Sigma_N(z) &= \kappa_{\mathcal{Z}}(z, z) - k_N(z) K_N^{\dagger} k_N(z)^* \\ &= \kappa_{\mathcal{Z}}(\cdot, z)^* \kappa_{\mathcal{Z}}(\cdot, z) \\ &\quad - \kappa_{\mathcal{Z}}(\cdot, z)^* \Phi_N (\Phi_N^* \Phi_N)^{\dagger} (\kappa_{\mathcal{Z}}(\cdot, z)^* \Phi_N)^* \\ &= \kappa_{\mathcal{Z}}(\cdot, z)^* (I - \Phi_N (\Phi_N^* \Phi_N)^{\dagger} \Phi_N^*) \kappa_{\mathcal{Z}}(\cdot, z) \\ &= \kappa_{\mathcal{Z}}(\cdot, z)^* (I - \Pi_N) \kappa_{\mathcal{Z}}(\cdot, z). \end{aligned}$$

Hence, for any  $v \in \mathbb{R}^p$ ,

$$\begin{aligned} v^T \Sigma_N(z) v &= v^T \kappa_{\mathcal{Z}}(\cdot, z)^* (I - \Pi_N) \kappa_{\mathcal{Z}}(\cdot, z) v \\ &= \|(I - \Pi_N) \kappa_{\mathcal{Z}}(\cdot, z) v\|_{\mathcal{H}_{\mathcal{Z}}}. \end{aligned}$$

Hence,  $\Sigma_N(z)$  is positive semi-definite. In particular,  $\Sigma_N(z) = 0$  iff  $(I - \Pi_N) \kappa_{\mathcal{Z}}(\cdot, z) v = 0$  for all  $v \in \mathbb{R}^p$ , that is, if  $\kappa_{\mathcal{Z}}(\cdot, z) v \in \mathcal{H}_N = \text{range}(\Phi_N)$ .

### C. Proof of Lemma 5

Since  $f_{N+1}$  interpolates  $\{(z_i, y_i)\}_{i=0}^{N-1}$ , we have for all  $v \in \mathcal{Y}$  and  $i = 0, \dots, N-1$ ,

$$\langle f_{N+1}, \kappa_{\mathcal{Z}}(\cdot, z_i)v \rangle_{\mathcal{H}_{\mathcal{Z}}} = \langle f_{N+1}(z_i), v \rangle_{\mathcal{Y}} = \langle y_i, v \rangle_{\mathcal{Y}}.$$

On the other hand, for  $i = 0, \dots, N-1$  and all  $v \in \mathcal{Y}$ ,

$$\begin{aligned} \langle f_{N+1}, \kappa_{\mathcal{Z}}(\cdot, z_i)v \rangle_{\mathcal{H}_{\mathcal{Z}}} &= \langle \Pi_N[f_{N+1}] + \Pi_N^{\perp}[f_{N+1}], \kappa_{\mathcal{Z}}(\cdot, z_i)v \rangle_{\mathcal{H}_{\mathcal{Z}}} \\ &= \langle \Pi_N[f_{N+1}], \kappa_{\mathcal{Z}}(\cdot, z_i)v \rangle_{\mathcal{H}_{\mathcal{Z}}}, \end{aligned}$$

where  $\Pi_N^{\perp} = I - \Pi_N$  and  $\Pi_N$  is the orthogonal projection onto  $\mathcal{H}_N$ . Since  $v$  is arbitrary,  $\Pi_N[f_{N+1}](z_i) = y_i$  for  $i = 0, \dots, N-1$ , i.e., both  $f_N$  and  $\Pi_N[f_{N+1}]$  interpolate  $\{(z_i, y_i)\}_{i=0}^{N-1}$ . Since  $f_N$  is the unique minimum-norm solution in  $\mathcal{H}_N$ , it must be the case that  $f_N = \Pi_N[f_{N+1}]$ . Hence,

$$f_{N+1} - f_N = f_{N+1} - \Pi_N[f_{N+1}] \in \mathcal{H}_N^{\perp}.$$

That is, there exists some  $\xi_N \in \mathcal{Y}$  such that

$$f_{N+1} - f_N = \Pi_N^{\perp}[\kappa_{\mathcal{Z}}(\cdot, z_N)\xi_N]. \quad (29)$$

We now determine  $\xi_N$ . By the reproducing property, we have

$$\begin{aligned} \langle f_{N+1}(z_N), v \rangle_{\mathcal{Y}} &= \langle f_{N+1}, \kappa_{\mathcal{Z}}(\cdot, z_N)v \rangle_{\mathcal{H}_{\mathcal{Z}}} \\ &= \langle f_N + \Pi_N^{\perp}[\kappa_{\mathcal{Z}}(\cdot, z_N)\xi_N], \kappa_{\mathcal{Z}}(\cdot, z_N)v \rangle_{\mathcal{H}_{\mathcal{Z}}} \\ &= \langle f_N, \kappa_{\mathcal{Z}}(\cdot, z_N)v \rangle_{\mathcal{H}_{\mathcal{Z}}} \\ &\quad + \langle \Pi_N^{\perp}[\kappa_{\mathcal{Z}}(\cdot, z_N)\xi_N], \kappa_{\mathcal{Z}}(\cdot, z_N)v \rangle_{\mathcal{H}_{\mathcal{Z}}}. \end{aligned} \quad (30)$$

For the last line, applying the reproducing property again, we can write the first term as  $\langle f_N, \kappa_{\mathcal{Z}}(\cdot, z_N)v \rangle_{\mathcal{H}_{\mathcal{Z}}} = \langle f_N(z_N), v \rangle_{\mathcal{Y}}$ . For the second term, by orthogonality we have

$$\begin{aligned} & \langle \Pi_N^\perp[\kappa_{\mathcal{Z}}(\cdot, z_N)\xi_N], \kappa_{\mathcal{Z}}(\cdot, z_N)v \rangle_{\mathcal{H}_{\mathcal{Z}}} \\ &= \langle \Pi_N^\perp[\kappa_{\mathcal{Z}}(\cdot, z_N)\xi_t], \Pi_N^\perp[\kappa_{\mathcal{Z}}(\cdot, z_N)v] \rangle_{\mathcal{H}_{\mathcal{Z}}} \\ &= \langle \xi_N, \kappa_{\mathcal{Z}}(z_N, z_N)v \rangle_{\mathcal{H}_{\mathcal{Z}}} \\ & \quad - \langle \Pi_N[\kappa_{\mathcal{Z}}(\cdot, z_N)\xi_N], \Pi_N[\kappa_{\mathcal{Z}}(\cdot, z_N)v] \rangle_{\mathcal{H}_{\mathcal{Z}}}, \end{aligned}$$

where we have used the fact that, since  $\Pi_N$  is an orthogonal projection,

$$\langle \Pi_N h, \Pi_N h' \rangle_{\mathcal{H}_{\mathcal{Z}}} = \langle h, \Pi_N h' \rangle_{\mathcal{H}_{\mathcal{Z}}} = \langle \Pi_N h, h' \rangle_{\mathcal{H}_{\mathcal{Z}}}$$

for all  $h, h' \in \mathcal{H}_{\mathcal{Z}}$ . Since  $\Pi_N$  is an orthogonal projection onto  $\mathcal{H}_N$ , we have

$$\Pi_N[\kappa_{\mathcal{Z}}(\cdot, z_N)\xi_N] = \sum_{j=0}^{N-1} \kappa_{\mathcal{Z}}(\cdot, z_j)\alpha_j,$$

and

$$\Pi_N[\kappa_{\mathcal{Z}}(\cdot, z_N)v] = \sum_{j=0}^{N-1} \kappa_{\mathcal{Z}}(\cdot, z_j)\beta_j,$$

where  $\bar{\alpha} = (\alpha_0, \dots, \alpha_{N-1}) \in \mathcal{Y}^{N-1}$  and  $\bar{\beta} = (\beta_0, \dots, \beta_{N-1}) \in \mathcal{Y}^{N-1}$  are the solutions of  $K_N \bar{\alpha} = k_N(z_N)\xi_N$  and  $K_N \bar{\beta} = k_N(z_N)v$ . Therefore,

$$\begin{aligned} & \langle \Pi_N[\kappa_{\mathcal{Z}}(\cdot, z_N)\xi_N], \Pi_N[\kappa_{\mathcal{Z}}(\cdot, z_N)v] \rangle_{\mathcal{H}_{\mathcal{Z}}} \\ &= \left\langle \sum_{j=0}^{N-1} \kappa_{\mathcal{Z}}(\cdot, z_j)\alpha_j, \sum_{j=0}^{N-1} \kappa_{\mathcal{Z}}(\cdot, z_j)\beta_j \right\rangle_{\mathcal{H}_{\mathcal{Z}}} \\ &= \langle \bar{\alpha}, K_N \bar{\beta} \rangle_{\mathcal{Y}^{\oplus(N-1)}} \\ &= \langle K_N^\dagger k_N(z_N)\xi_N, K_N^\dagger k_N(z_N)v \rangle_{\mathcal{Y}^{\oplus(N-1)}} \\ &= \langle \xi_N, k_N^*(z_N)K_N^\dagger k_N(z_N)v \rangle_{\mathcal{Y}}, \end{aligned}$$

where the last line holds since  $K_N^\dagger K_N K_N^\dagger = K_N^\dagger$ .

Putting everything together and using the definition of  $\Sigma_N(z_N)$ , we arrive at

$$\begin{aligned} & \langle \Pi_N^\perp[\kappa_{\mathcal{Z}}(\cdot, z_N)\xi_N], \Pi_N^\perp[\kappa_{\mathcal{Z}}(\cdot, z_N)v] \rangle_{\mathcal{H}_{\mathcal{Z}}} \\ &= \langle \xi_N, \kappa_{\mathcal{Z}}(z_N, z_N)v \rangle_{\mathcal{Y}} - \langle \xi_N, k_N^*(z_N)K_N^\dagger k_N(z_N)v \rangle_{\mathcal{Y}} \\ &= \langle \xi_N, \Sigma_N(z_N)v \rangle_{\mathcal{Y}}. \end{aligned}$$

Plugging the above relation back into (30) and using the fact that  $v \in \mathcal{Y}$  is arbitrary and that  $\Sigma_N(z_N)$  is self-adjoint, we see that  $\xi_N$  is determined by the relation

$$f_{N+1}(z_N) = f_N(z_N) + \Sigma_N(z_N)\xi_N.$$

As  $f_{N+1}$  interpolates  $(z_N, y_{N+})$ , i.e.,  $y_{N+} = f_{N+1}(z_N)$ , we can further write

$$\Sigma_N(z_N)\xi_N = y_{N+} - f_N(z_N). \quad (31)$$

Hence, when  $\Sigma_N(z_N) = 0$ , we will have  $f_N(z_N) = f_{N+1}(z_N) = y_{N+}$ .

Next, applying the Pythagorean theorem in (29), we have

$$\begin{aligned} \|f_{N+1}\|_{\mathcal{H}_{\mathcal{Z}}}^2 &= \|f_N\|_{\mathcal{H}_{\mathcal{Z}}}^2 + \|\Pi_N^\perp[\kappa_{\mathcal{Z}}(\cdot, z_N)\xi_t]\|_{\mathcal{H}_{\mathcal{Z}}}^2 \\ &= \|f_N\|_{\mathcal{H}_{\mathcal{Z}}}^2 + \langle \xi_N, \Sigma_N(z_N)\xi_N \rangle_{\mathcal{Y}}, \end{aligned}$$

where the last line follows from the definition of  $\Sigma_N$ . When  $\Sigma_N(z_N) \succ 0$ , from (31) we have

$$\begin{aligned} \|f_{N+1}\|_{\mathcal{H}_{\mathcal{Z}}}^2 - \|f_N\|_{\mathcal{H}_{\mathcal{Z}}}^2 &= \langle \Sigma_N^{-1}(z_N)(y_{N+} - f_N(z_N)), y_{N+} - f_N(z_N) \rangle_{\mathcal{Y}} \\ &= \left\| \Sigma_N^{-1/2}(z_N)(y_{N+} - f_N(z_N)) \right\|_{\mathcal{Y}}^2. \end{aligned}$$

This concludes the proof.

## APPENDIX C

### PROOFS FOR RESULTS IN SECTION V

#### A. Proof of Theorem 9

In our model (22), we can view  $v = \phi(u)$  as the  $q$ -dimensional input to the LTI system parametrized by  $(A, B, C, D)$ . Consequently, the argument in the proof in [26, Theorem 1] extends directly to our setting and guarantees that the matrix

$$\mathcal{M} := \begin{bmatrix} U_p^\phi \\ U_f^\phi \\ X_0 \end{bmatrix} \in \mathbb{R}^{(2Lq+n) \times (T-2L+1)}$$

has full row rank.

( $\Rightarrow$ ): Suppose that  $\begin{bmatrix} \Phi_{0:2L-1} \\ y_{0:2L-1} \end{bmatrix}$  is a valid input-output trajectory of (22). Then there exists an initial state  $\hat{x}_0$  such that

$$y_{0:L-1} = \mathcal{O}_L \hat{x}_0 + \mathcal{T}_L \Phi_{0:L-1}.$$

Since  $\mathcal{M}$  has full row rank, for any  $\begin{bmatrix} \Phi_{0:L-1} \\ \Phi_{L:2L-1} \\ \hat{x}_0 \end{bmatrix} \in \mathbb{R}^{2Lq} \oplus \mathbb{R}^n$ ,

there exists some  $\xi \in \mathbb{R}^{T-2L+1}$  such that

$$\begin{bmatrix} \Phi_{0:L-1} \\ \Phi_{L:2L-1} \\ \hat{x}_0 \end{bmatrix} = \mathcal{M}\xi.$$

Thus, we have

$$\begin{aligned} \begin{bmatrix} \Phi_{0:L-1} \\ y_{0:L-1} \end{bmatrix} &= \begin{bmatrix} I & 0 \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} \Phi_{0:L-1} \\ \hat{x}_0 \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} U_p^\phi \\ X_0 \end{bmatrix} \xi \\ &= \begin{bmatrix} U_p^\phi \\ Y_p \end{bmatrix} \xi. \end{aligned}$$

Moreover, from the dynamics (22), we have

$$X_f = A^L X_0 + \Delta_L U_p^\phi. \quad (32)$$

Hence, at time  $t = L$ ,

$$\begin{aligned} \hat{x}_L &= A^L \hat{x}_0 + \Delta_L \Phi_{0:L-1} \\ &= (A^L X_0 + \Delta_L U_p^\phi) \xi \\ &= X_f \xi. \end{aligned}$$

Thus, starting from  $\hat{x}_L = X_f \xi$  with input sequence  $\Phi_{L:2L-1} = U_f^\phi \xi$ , we have

$$\begin{aligned} \begin{bmatrix} \Phi_{L:2L-1} \\ y_{L:2L-1} \end{bmatrix} &= \begin{bmatrix} I & 0 \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} \Phi_{L:2L-1} \\ \hat{x}_L \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} U_f^\phi \\ X_f \end{bmatrix} \xi \\ &= \begin{bmatrix} U_f^\phi \\ Y_f \end{bmatrix} \xi. \end{aligned}$$

( $\Leftarrow$ ) Since  $\begin{bmatrix} u_{0:L-1} \\ y_{0:L-1} \end{bmatrix} = \begin{bmatrix} U_p^\phi \\ Y_p \end{bmatrix} \xi$ , and  $\begin{bmatrix} U_p^\phi \\ Y_p \end{bmatrix}$  is constrained by (22), we have

$$\begin{bmatrix} \Phi_{0:L-1} \\ y_{0:L-1} \end{bmatrix} = \begin{bmatrix} U_p^\phi \\ Y_p \end{bmatrix} \xi = \begin{bmatrix} I & 0 \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} U_p^\phi \\ X_0 \end{bmatrix} \xi.$$

Hence,

$$y_{0:L-1} = Y_p \xi = \mathcal{O}_L X_0 \xi + \mathcal{T}_L U_p^\phi \xi$$

Since  $\Phi_{0:L-1} = U_p^\phi \xi$ , we conclude that  $y_{0:L-1}$  is the output from  $\hat{x}_0 := X_0 \xi$  with input  $u_{0:L-1}$ . As  $(\Phi_{0:L-1}, y_{0:L-1})$  is a valid input-output trajectory, we can use (22) and  $X_f$  defined in (32) to obtain the state vector at time  $L$  as

$$\hat{x}_L = A^L \hat{x}_0 + \Delta_L \Phi_{0:L-1} = A^L X_0 \xi + \Delta_L U_p^\phi \xi = X_f \xi. \quad (33)$$

For the second segment, using the assumptions and the matrix input-output relation for  $U_f^\phi, Y_f$ , we have

$$\begin{bmatrix} \Phi_{L:2L-1} \\ y_{L:2L-1} \end{bmatrix} = \begin{bmatrix} U_f^\phi \\ Y_f \end{bmatrix} \xi = \begin{bmatrix} I & 0 \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} U_f^\phi \\ X_f \end{bmatrix} \xi.$$

Thus, we have

$$\begin{aligned} y_{L:2L-1} &= Y_f \xi \\ &= \mathcal{O}_L X_f \xi + \mathcal{T}_L U_f^\phi \xi \\ &= \mathcal{O}_L \hat{x}_L + \mathcal{T}_L \Phi_{L:2L-1}, \end{aligned}$$

where the last equality follows from (33). Therefore, we conclude that  $y_{L:2L-1}$  is the output from  $\hat{x}_L = X_f \xi$  with input  $\Phi_{L:2L-1}$ .

Putting everything together, we have

$$\begin{bmatrix} \Phi_{0:L-1} \\ y_{0:L-1} \end{bmatrix} = \begin{bmatrix} U_p^\phi \\ Y_p \end{bmatrix} \xi, \quad \begin{bmatrix} \Phi_{L:2L-1} \\ y_{L:2L-1} \end{bmatrix} = \begin{bmatrix} U_f^\phi \\ Y_f \end{bmatrix} \xi.$$

Multiplying both sides of the first equation by  $\begin{bmatrix} U_p^{\phi*} & 0 \\ 0 & Y_p^* \end{bmatrix}$

and the second equation by  $\begin{bmatrix} U_f^{\phi*} & 0 \\ 0 & Y_f^* \end{bmatrix}$  proves the claim.

## REFERENCES

- [1] J. C. Willems, "From time series to linear system—part i. finite dimensional linear time invariant systems," *Automatica*, vol. 22, no. 5, pp. 561–580, 1986.
- [2] —, "Paradigms and puzzles in the theory of dynamical systems," *IEEE Transactions on Automatic Control*, vol. 36, no. 3, pp. 259–294, 1991.
- [3] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. De Moor, "A note on persistency of excitation," *Systems & Control Letters*, vol. 54, no. 4, pp. 325–329, 2005.
- [4] I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor, *Exact and approximate modeling of linear systems: A behavioral approach*. SIAM, 2006.
- [5] I. Markovsky and P. Rapisarda, "Data-driven simulation and control," *International Journal of Control*, vol. 81, no. 12, pp. 1946–1959, 2008.
- [6] J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control: In the shallows of the deep," in *2019 18th European control conference (ECC)*. IEEE, 2019, pp. 307–312.
- [7] F. Dörfler, J. Coulson, and I. Markovsky, "Bridging direct and indirect data-driven control formulations via regularizations and relaxations," *IEEE Transactions on Automatic Control*, vol. 68, no. 2, pp. 883–897, 2022.
- [8] H. J. Van Waarde, J. Eising, M. K. Camlibel, and H. L. Trentelman, "The informativity approach to data-driven analysis and control," *IEEE Control Systems Magazine*, vol. 43, no. 6, pp. 32–66, 2023.
- [9] I. Markovsky and F. Dörfler, "Data-driven dynamic interpolation and approximation," *Automatica*, vol. 135, p. 110008, 2022.
- [10] J. G. Rueda-Escobedo and J. Schiffer, "Data-driven internal model control of second-order discrete volterra systems," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 4572–4579.
- [11] M. Alsalti, J. Berberich, V. G. Lopez, F. Allgöwer, and M. A. Müller, "Data-based system analysis and control of flat nonlinear systems," in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 1484–1489.
- [12] J. Berberich and F. Allgöwer, "A trajectory-based framework for data-driven system analysis and control," in *2020 European Control Conference (ECC)*. IEEE, 2020, pp. 1365–1370.
- [13] X. Shang, J. Cortés, and Y. Zheng, "Willems' fundamental lemma for nonlinear systems with koopman linear embedding," *IEEE Control Systems Letters*, 2024.
- [14] L. Huang, J. Lygeros, and F. Dörfler, "Robust and kernelized data-enabled predictive control for nonlinear systems," *IEEE Transactions on Control Systems Technology*, vol. 32, no. 2, pp. 611–624, 2023.
- [15] O. Molodchyk and T. Faulwasser, "Exploring the links between the fundamental lemma and kernel regression," *IEEE Control Systems Letters*, vol. 8, pp. 2045–2050, 2024.
- [16] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *International conference on computational learning theory*. Springer, 2001, pp. 416–426.
- [17] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: Theory—Implementation—Applications*. Kluwer, 1996.
- [18] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.
- [19] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American mathematical society*, vol. 68, no. 3, pp. 337–404, 1950.
- [20] C. Carmeli, E. De Vito, and A. Toigo, "Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem," *Analysis and Applications*, vol. 04, no. 04, pp. 377–408, Oct. 2006.
- [21] J. Moore, "Persistence of excitation in extended least squares," *IEEE Transactions on Automatic Control*, vol. 28, no. 1, pp. 60–68, 1983.
- [22] M. Green and J. B. Moore, "Persistence of excitation in linear systems," *Systems & Control Letters*, vol. 7, no. 5, pp. 351–360, 1986.
- [23] R. De Figueiredo and T. Dwyer, "A best approximation framework and implementation for simulation of large-scale nonlinear systems," *IEEE Transactions on circuits and systems*, vol. 27, no. 11, pp. 1005–1014, 1980.

- [24] C. A. Micchelli and M. Pontil, “On learning vector-valued functions,” *Neural computation*, vol. 17, no. 1, pp. 177–204, 2005.
- [25] T. Liang and B. Recht, “Interpolating classifiers make few mistakes,” *Journal of Machine Learning Research*, vol. 24, no. 20, pp. 1–27, 2023.
- [26] H. J. Van Waarde, C. De Persis, M. K. Camlibel, and P. Tesi, “Willems’ fundamental lemma for state-space systems and its extension to multiple datasets,” *IEEE Control Systems Letters*, vol. 4, no. 3, pp. 602–607, 2020.