A Behavioral Perspective on Subspace and Interpolation Methods for Simulation of Linear Time-Invariant Systems

Boya Hou[1] and Maxim Raginsky[2]

***Abstract*— In this paper, we revisit Jan Willems' behavioral approach to data-driven simulation of control systems from several complementary perspectives. Starting from an input/output model of a linear time-invariant system, we elucidate the links between data-driven reconstruction of the model behavior and least-squares prediction of the system response to a specified input given an initial condition and a previously measured input/output trajectory. We next establish the connection between the behavioral approach and subspace identification via oblique projection by providing a sufficient condition for the validity of subspace identification via persistence of excitation. When the memory length (or lag) of the input/output model is significantly larger than the minimal state dimension, the state vector can encode the system's memory more efficiently. Finally, we link the behavioral approach to minimum-norm interpolation with vector-valued observations. These results provide a unified perspective on the behavioral approach, subspace identification, and minimum-norm interpolation.**

I. Introduction

Mathematical study of dynamical systems aims at describing how a given system evolves over time subject to its laws, initial conditions, and inputs (if any). This evolution can be understood internally using state-space methods or externally through the system's behavior. Pioneered by Jan Willems in a series of papers starting with [1] (see [2] for an overview), the behavioral approach provides a nonparametric representation of dynamical systems by identifying them with sets of trajectories. In the context of linear time-invariant systems, a central result of the behavioral approach is the so-called *fundamental lemma* [3], which states that the set of all possible trajectories that could be generated by a controllable linear time-invariant system in a finite time interval can be reconstructed from a single trajectory driven by a persistently exciting input. This result has played a foundational role in the recent advances of data-driven control; see, e.g., [4, 5, 6, 7, 8] and references therein.

The philosophy underlying the data-driven approach, as articulated in a clearly written paper of Markovsky and Rapisarda [5], is that solutions to problems of system simulation or control should be constructed without an explicit system identification step. For instance, in the context of system simulation the procedure that takes the "online" data (initial condition and a subsequent input of interest) and the "offline" data (previous measurements of the system behavior) and predicts the resulting output should not rely on identifying the system transfer function, impulse response, or state-space model. Instead, one aims for a direct approach that views simulation as an instance of a "missing data" problem pertaining to the system behavior.

In this paper, we revisit several classical approaches to system modeling, namely subspace predictive methods based on least squares [9], subspace identification [10], and minimum-norm interpolation [11], and reinterpret them from the perspective of data-driven system modeling informed by the behavioral view. The common theme underlying all these methods is that, like in the newer data-driven schemes and like in the fundamental lemma of Willems, suitable linear combinations of "offline" data are sufficient for making predictions regarding "online" data.

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 $A^\dagger$. We will use $\|\cdot\|_2$ to denote the Euclidean (ℓ^2) norm on vectors and $\|\cdot\|_{\mathrm{F}}$ to denote the Frobenius (or Hilbert–Schmidt) norm on matrices. 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 rowspace of A on the rowspace of C along the rowspace of B is defined as

$$A \underset{B}{/} C := A \begin{bmatrix} C^\intercal & B^\intercal \end{bmatrix} \left(\begin{bmatrix} CC^\intercal & CB^\intercal \\ BC^\intercal & BB^\intercal \end{bmatrix}^\dagger \right)_{\text{first } r \text{ columns}} C \quad (1)$$

(see [10, 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

$$H_L(w_{0:T-1}) := \begin{bmatrix} w_0 & w_1 & \dots & w_{T-L} \\ w_1 & w_2 & \dots & w_{T-L+1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{L-1} & w_L & \dots & w_{T-1} \end{bmatrix}. \quad (2)$$

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) $\{w_t\}_{t\in\mathbb{Z}_+}$, σ is the backward shift operator defined by $(\sigma w)_t := w_{t+1}$.

II. 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 sequence space $(\mathbb{R}^q)^{\mathbb{Z}_+}$ which is shift-invariant, i.e., $\sigma\mathcal{B} \subseteq \mathcal{B}$.

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

[1] 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

[2] 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

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 $u : \mathbb{Z}_+ \to \mathbb{R}^m$ and an output trajectory $y : \mathbb{Z}_+ \to \mathbb{R}^m$ as $w = \begin{bmatrix} u \\ y \end{bmatrix}$, such that the input is free in the sense that

$$\left\{ u : \mathbb{Z}_+ \to \mathbb{R}^m \,\middle|\, \begin{bmatrix} u \\ y \end{bmatrix} \in \mathcal{B} \text{ for some } y : \mathbb{Z}_+ \to \mathbb{R}^p \right\} \simeq (\mathbb{R}^m)^{\mathbb{Z}_+} \tag{3}$$

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

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. In the following sections, we will examine some existing data-driven methods from the behavioral perspective and highlight their similarities and differences with the assumptions underlying Lemma 1.

III. Connection to least-squares prediction

We begin by considering a discrete-time linear system in input/output form, a specific type of an autoregressive model. This model relates the sequence of $\mathbb{R}^m$-valued inputs $\{u_t\}_{t \in \mathbb{Z}_+}$ to the sequence of $\mathbb{R}^p$-valued outputs $\{y_t\}_{t \in \mathbb{Z}_+}$ through

$$y_{t+L} + \sum_{k=1}^{L} A_k y_{t+L-k} = \sum_{l=0}^{L} B_l u_{t+L-l}, \qquad t \in \mathbb{Z}_+ \tag{4}$$

where L is the system lag and where $A_1, \ldots, A_L \in \mathbb{R}^{p \times p}$ and $B_0, \ldots, B_L \in \mathbb{R}^{p \times m}$ are given matrices. The specification of $\{u_i\}_{i=0}^{L}$ and $\{y_i\}_{i=0}^{L-1}$ provides the initial condition. We can express (4) more succinctly as follows. Define the polynomial matrices $P(\sigma) := \sum_{k=0}^{L} A_{L-k}\sigma^k$ (with $A_0 = I_p$) and $Q(\sigma) := \sum_{l=0}^{L} B_{L-l}\sigma^l$, where σ is the backward shift operator. Then the input/output model (4) defines a linear system with behavior

$$\mathcal{B}(P, Q) = \left\{ \begin{bmatrix} u \\ y \end{bmatrix} : \mathbb{Z}_+ \to \mathbb{R}^{m+p} \,\middle|\, P(\sigma)y = Q(\sigma)u \right\}. \tag{5}$$

The restriction of the behavior $\mathcal{B}(P, Q)$ to the time interval $\{t, \ldots, t+L-1\}$ is denoted by

$$\mathcal{B}(P, Q)|_{t:t+L-1} = \left\{ \begin{bmatrix} u_{t:t+L-1} \\ y_{t:t+L-1} \end{bmatrix} \,\middle|\, \begin{bmatrix} u \\ y \end{bmatrix} \in \mathcal{B}(P, Q) \right\}. \tag{6}$$

When $t = 0$, we will use $\mathcal{B}(P, Q)|_L$ as shorthand for $\mathcal{B}(P, Q)|_{t:t+L-1}$. We have the following simple lemma.

Lemma 2. *There exists a matrix $K_\star \in \mathbb{R}^{Lp \times L(2m+p)}$, such that each $\begin{bmatrix} u \\ y \end{bmatrix} \in \mathcal{B}(P, Q)$ satisfies*

$$y_{t+L:t+2L-1} = K_\star \begin{bmatrix} u_{t:t+L-1} \\ y_{t:t+L-1} \\ u_{t+L:t+2L-1} \end{bmatrix}, \quad t \in \mathbb{Z}_+. \tag{7}$$

Using this lemma, we can characterize (4) through the behavior parameterized by $K_\star$:

$$\mathcal{B}(K_\star) = \left\{ \begin{bmatrix} u \\ y \end{bmatrix} \,\middle|\, y_{t+L:t+2L-1} = K_\star \begin{bmatrix} u_{t:t+L-1} \\ y_{t:t+L-1} \\ u_{t+L:t+2L-1} \end{bmatrix},\ t \in \mathbb{Z}_+ \right\}. \tag{8}$$

The restriction of the behavior to length-$2L$ sequences is defined as

$$\mathcal{B}(K_\star)|_{2L} = \left\{ \begin{bmatrix} u_{0:2L-1} \\ y_{0:2L-1} \end{bmatrix} \,\middle|\, \exists \begin{bmatrix} \bar{u} \\ \bar{y} \end{bmatrix} \in \mathcal{B}(K_\star) \text{ s.t. } \begin{bmatrix} \bar{u}_{0:2L-1} \\ \bar{y}_{0:2L-1} \end{bmatrix} = \begin{bmatrix} u_{0:2L-1} \\ y_{0:2L-1} \end{bmatrix} \right\}. \tag{9}$$

We now consider the data-driven simulation problem. Let $\{(u^{\mathrm{d}}_t, y^{\mathrm{d}}_t)\}_{t=0}^{T-1}$ denote input/output data of length T collected from measurements of (4). Partition the depth-$2L$ Hankel matrices $H_{2L}(u^{\mathrm{d}}_{0:T-1})$ and $H_{2L}(y^{\mathrm{d}}_{0:T-1})$ as

$$\begin{aligned}
\begin{bmatrix} U_{\mathrm{p}} \\ \hline U_{\mathrm{f}} \end{bmatrix} &:= \begin{bmatrix} u^{\mathrm{d}}_0 & u^{\mathrm{d}}_1 & \cdots & u^{\mathrm{d}}_{T-2L} \\ \vdots & \vdots & \ddots & \vdots \\ u^{\mathrm{d}}_{L-1} & u^{\mathrm{d}}_L & \cdots & u^{\mathrm{d}}_{T-L-1} \\ \hline u^{\mathrm{d}}_L & u^{\mathrm{d}}_{L+1} & \cdots & u^{\mathrm{d}}_{T-L} \\ \vdots & \vdots & \ddots & \vdots \\ u^{\mathrm{d}}_{2L-1} & u^{\mathrm{d}}_{2L} & \cdots & u^{\mathrm{d}}_{T-1} \end{bmatrix}, \\
\begin{bmatrix} Y_{\mathrm{p}} \\ \hline Y_{\mathrm{f}} \end{bmatrix} &:= \begin{bmatrix} y^{\mathrm{d}}_0 & y^{\mathrm{d}}_1 & \cdots & y^{\mathrm{d}}_{T-2L} \\ \vdots & \vdots & \ddots & \vdots \\ y^{\mathrm{d}}_{L-1} & y^{\mathrm{d}}_L & \cdots & y^{\mathrm{d}}_{T-L-1} \\ \hline y^{\mathrm{d}}_L & y^{\mathrm{d}}_{L+1} & \cdots & y^{\mathrm{d}}_{T-L} \\ \vdots & \vdots & \ddots & \vdots \\ y^{\mathrm{d}}_{2L-1} & y^{\mathrm{d}}_{2L} & \cdots & y^{\mathrm{d}}_{T-1} \end{bmatrix},
\end{aligned} \tag{10}$$

where p and f designates a partition into "past" and "future" data. Let H_p denote the concatenation of U_p and Y_p, $H_\mathrm{p} := \begin{bmatrix} U_\mathrm{p} \\ Y_\mathrm{p} \end{bmatrix}$. Similarly, let $H_\mathrm{f} := \begin{bmatrix} U_\mathrm{f} \\ Y_\mathrm{f} \end{bmatrix}$. By Lemma 2, Y_f can be expressed as a linear combination of H_p and U_f as

$$Y_\mathrm{f} = K_\star \begin{bmatrix} H_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}. \tag{11}$$

This relation is the basis of subspace predictive control [9]. Let $\widehat{K} \in \mathbb{R}^{Lp \times L(2m+p)}$ be a solution of the following least-squares linear regression problem:

$$\widehat{K} := \operatorname*{argmin}_{K \in \mathbb{R}^{Lp \times L(2m+p)}} \left\| Y_\mathrm{f} - K \begin{bmatrix} H_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix} \right\|_\mathrm{F}^2; \tag{12}$$

that is, $\widehat{K} \begin{bmatrix} H_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}$ is the optimal linear predictor of the output Y_f given the past H_p and the input U_f. Define $G^u_\mathrm{pp} = U_\mathrm{p}^\top U_\mathrm{p}$, $G^y_\mathrm{pp} = Y_\mathrm{p}^\top Y_\mathrm{p}$, $G^u_\mathrm{ff} = U_\mathrm{f}^\top U_\mathrm{f}$, and let $\overline{G}_\mathrm{pf} := G^u_\mathrm{pp} + G^y_\mathrm{pp} + G^u_\mathrm{ff}$. The optimal (minimum norm) solution is given by

$$\begin{aligned} \widehat{K} &= Y_\mathrm{f} \begin{bmatrix} H_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}^\dagger \\ &= Y_\mathrm{f} \left(\begin{bmatrix} H_\mathrm{p}^\top & U_\mathrm{f}^\top \end{bmatrix} \begin{bmatrix} H_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix} \right)^\dagger \begin{bmatrix} H_\mathrm{p}^\top & U_\mathrm{f}^\top \end{bmatrix} \\ &= Y_\mathrm{f} \left(G^u_\mathrm{pp} + G^y_\mathrm{pp} + G^u_\mathrm{ff} \right)^\dagger \begin{bmatrix} H_\mathrm{p}^\top & U_\mathrm{f}^\top \end{bmatrix} \\ &= Y_\mathrm{f} \overline{G}_\mathrm{pf}{}^\dagger \begin{bmatrix} H_\mathrm{p}^\top & U_\mathrm{f}^\top \end{bmatrix}. \end{aligned} \tag{13}$$

When $\begin{bmatrix} H_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}$ has full row rank, we have $\widehat{K} = K_\star$ (i.e., the identification is exact), since

$$\widehat{K} = Y_\mathrm{f} \begin{bmatrix} H_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}^\dagger = \left(K_\star \begin{bmatrix} H_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix} \right) \begin{bmatrix} H_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}^\dagger = K_\star. \tag{14}$$

We next present a method for computing the output of a system given a specific input and initial condition. This was referred to as the data-driven simulation problem in [5].

Theorem 3. *Let $\widehat{K}$ be a solution of* (12) *computed from the data $\{(u_t^\mathrm{d}, y_t^\mathrm{d})\}_{t=0}^{T-1}$. Given an initial length-L input-output trajectory $(u_{0:L-1}, y_{0:L-1})$ and a subsequent length-L input $u_{L:2L-1}$, construct the predictor $\widehat{y}_{L:2L-1}$ by*

$$\widehat{y}_{L:2L-1} = \widehat{K} \begin{bmatrix} u_{0:L-1} \\ y_{0:L-1} \\ u_{L:2L-1} \end{bmatrix}. \tag{15}$$

Suppose that

$$\begin{bmatrix} u_{0:L-1} \\ y_{0:L-1} \\ u_{L:2L-1} \end{bmatrix} \in \operatorname{colspace} \left(\begin{bmatrix} U_\mathrm{p} \\ Y_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix} \right). \tag{16}$$

Then we have

$$\begin{bmatrix} u_{0:L-1} \\ y_{0:L-1} \end{bmatrix} \wedge \begin{bmatrix} u_{L:2L-1} \\ \widehat{y}_{L:2L-1} \end{bmatrix} \in \mathcal{B}(K_\star) \mid_{2L}, \tag{17}$$

where $\wedge$ denotes concatenation of trajectories.

In the case where the conditions of Theorem 3 hold, we can first compute $\widehat{K}$ offline (i.e., from the observed data), then use it to predict $\widehat{y}_{L:2L-1}$ via (15), and the prediction is guaranteed to solve the simulation problem. In order to apply this method, we need to know the system lag L. On the other hand, the fundamental lemma states that, if the system is controllable and if the input data $\{u_t^\mathrm{d}\}_{t=0}^{T-1}$ satisfy the persistency of excitation condition of order $2L+n$ with n being the dimension of any minimal state-space realization of the system, then $\{u_i, y_i\}_{i=0}^{2L-1}$ is a valid length-$2L$ input-output trajectory of (4) if and only if there exists some vector $\xi \in \mathbb{R}^{T-2L+1}$, such that

$$\begin{bmatrix} u_{0:2L-1} \\ y_{0:2L-1} \end{bmatrix} = \begin{bmatrix} U_\mathrm{p} \\ Y_\mathrm{p} \\ U_\mathrm{f} \\ Y_\mathrm{f} \end{bmatrix} \xi. \tag{18}$$

In this case, the predictor simplifies to

$$\widehat{y}_{L:2L-1} = \widehat{K} \begin{bmatrix} u_{0:L-1} \\ y_{0:L-1} \\ u_{L:2L-1} \end{bmatrix} = \widehat{K} \begin{bmatrix} U_\mathrm{p} \\ Y_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix} \xi = Y_\mathrm{f} \xi. \tag{19}$$

IV. Connection to Subspace Identification

We now consider the setting when the input/output model (4) admits a state-space realization of the form

$$\begin{aligned} x_{t+1} &= A x_t + B u_t, \\ y_t &= C x_t + D u_t, \end{aligned} \tag{20}$$

where $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^m$ is the control input, $y \in \mathbb{R}^p$ is the output, and $A \in \mathbb{R}^{m \times m}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, $D \in \mathbb{R}^{p \times m}$ are constant matrices. One trivial way to construct the state vector is by setting $x_t = [u_{t-L}^\top, \cdots, u_t^\top, y_{t-L}^\top, \cdots y_{t-1}^\top]^\top$, which requires specifying the initial condition $x_0 = [u_{-L}^\top, \ldots, u_0^\top, y_{-L}^\top, \ldots, y_{-1}^\top]^\top$.

Recall the definitions of the L-step extended observability matrix $\mathcal{O}_L$, L-step controllability matrix $\mathcal{C}_L$, and the reversed L-step controllability matrix Δ_L associated with (20):

$$\begin{aligned} \mathcal{O}_L &:= \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{bmatrix}, \\ \mathcal{C}_L &:= \begin{bmatrix} B & AB & A^2B & \cdots & A^{L-1}B \end{bmatrix}, \\ \Delta_L &:= \begin{bmatrix} A^{L-1}B & \cdots & AB & B \end{bmatrix}. \end{aligned} \tag{21}$$

We also have the block Toeplitz matrix

$$\mathcal{T}_L := \begin{bmatrix} D & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ CB & D & \mathbf{0} & \cdots & \mathbf{0} \\ CAB & CB & D & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{L-1}B & CA^{L-2}B & \cdots & CB & D \end{bmatrix}. \tag{22}$$

(See, e.g., Section 2.1 of [10].) Consider again input-output data $\{(u_t^\mathrm{d}, y_t^\mathrm{d})\}_{t=0}^{T-1}$, which we assume to have been generated by (20) starting from some initial condition $x_0 \in \mathbb{R}^n$. Then we have the matrix input-output equations

$$Y_\mathrm{p} = \mathcal{O}_L X_0 + \mathcal{T}_L U_\mathrm{p}, \quad Y_\mathrm{f} = \mathcal{O}_L X_\mathrm{f} + \mathcal{T}_L U_\mathrm{f}, \tag{23}$$

where $X_0 := \begin{bmatrix} x_0 & \cdots & x_{T-2L} \end{bmatrix} \in \mathbb{R}^{n \times (T-2L+1)}$ and $X_\mathrm{f} := \begin{bmatrix} x_L & \cdots & x_{T-L} \end{bmatrix} \in \mathbb{R}^{n \times (T-2L+1)}$.

The idea behind subspace identification is that, under certain regularity conditions, the state matrix X_f and the observability matrix $\mathcal{O}_L$ can be reconstructed directly from the input/output data without knowledge of the system matrices $A, B, C.D$. Specifically, let Π denote the oblique projection of the rowspace of Y_f onto the rowspace of H_p along the rowspace of U_f: $\Pi := Y_\mathrm{f} \,/_{U_\mathrm{f}}\, H_\mathrm{p}$. Let the singular value decomposition of Π be given by

$$\Pi = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \Sigma_1 V_1^T, \tag{24}$$

where $\Sigma_1 \in \mathbb{R}^{r \times r}$ with $r = \text{rank}\ (\Pi)$. Then, up to a similarity transformation, we have $X_\mathrm{f} = \Sigma_1^{1/2} V_1^T$ (see, e.g., [10, Theorem 2, Chapter 2]). However, the underlying regularity conditions involve the state and therefore cannot be verified on the basis of input/output data alone. The following result shows that the assumptions required to perform oblique projection are satisfied under a suitable PE condition on the input data:

Theorem 4. *Let $\left\{(u_t^\mathrm{d}, y_t^\mathrm{d})\right\}_{t=0}^{T-1}$ be an input/output sequence of length T generated by* (20)*. Suppose that the state-space realization in* (20) *is minimal (controllable and observable) and that $\mathrm{rank}(\mathcal{O}_L) = n$. Suppose further that the input sequence is persistently exciting of order $2L + n$, i.e., $\mathrm{rank}(H_{2L+n}(u_{0:T-1})) = (2L + n)m$. Then, $\Pi = \mathcal{O}_L X_\mathrm{f}$ and $X_\mathrm{f} = \Sigma_1^{1/2} V_1^T$.*

The next result explicitly connects subspace identification to the fundamental lemma. As in the setting of Theorem 3, we partition a length-$2L$ sequence into two length-L segments. The sequence constitutes a valid length-$2L$ trajectory of (20) if there exist initial conditions consistent with both segments.

Theorem 5. *Let (A, B) be controllable and (A, C) observable. Let input-output data $\{(u_t^\mathrm{d}, y_t^\mathrm{d})\}_{t=0}^{T-1}$ of length T be given, and suppose the input is persistently exciting of order $2L + n$, where 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 trajectory of system* (20) *if and only if there exists some vector $\xi \in \mathbb{R}^{T-2L+1}$ such that*

$$\begin{bmatrix} u_{0:L-1} \\ y_{0:L-1} \end{bmatrix} = \begin{bmatrix} U_\mathrm{p} \\ Y_\mathrm{p} \end{bmatrix} \xi \text{ and } \begin{bmatrix} u_{L:2L-1} \\ y_{L:2L-1} \end{bmatrix} = \begin{bmatrix} U_\mathrm{f} \\ Y_\mathrm{f} \end{bmatrix} \xi. \tag{25}$$

In the context of the above theorem, we can view $(U_\mathrm{p}, Y_\mathrm{p}, U_\mathrm{f}, Y_\mathrm{f})$ as "offline" training data generated by (20). By splitting the "online" testing sequence into two parts (past and future), we test whether one can interpolate the past and future using the same coefficient, where the initial length-L segment of the testing sequence specifies the initial condition for the subsequent length-L segment. A similar observation on the specification of initial condition has been given in [5, Proposition 1]. However, in the proof of Theorem 5 we further show that the trajectory passes through a state $\widehat{x}_L$ at time L that is a linear combination of states from the training data, i.e., $\widehat{x}_L = X_\mathrm{f}\xi$ for some ξ.

Finally, we revisit the least-squares prediction perspective of Theorem 3. Notice that we did not require the system to be controllable. Instead, we assumed the Gram matrix $\overline{G}_\mathrm{pf}$ to be invertible. Leveraging the link with subspace identification, we establish in the following lemma that, under the assumption that the system is controllable and that the input is persistently exciting of order $2L + n$, the Gram matrix has full rank $2Lm + n$. We further remark that the input Hankel matrix has more columns than rows, i.e., $T - 2L + 1 \geq 2Lm$. Hence, when $\overline{G}_\mathrm{pf}$ has full rank, i.e., $T - 2L + 1 = 2Lm + n$, then it is invertible.

Lemma 6. *Suppose (A, B) is controllable and (A, C) observable. In addition, suppose that rank $(\mathcal{O}_L) = n$ and the input is persistently exciting of order $2L + n$. Then, rank $(\overline{G}_\mathrm{pf}) = 2Lm + n$.*

V. Connection to minimum norm interpolation

Consider again the input-output model (4), and define $z_t := [u_{t-L}^T, \cdots, u_t^T, y_{t-L}^T, \cdots y_{t-1}^T]^T$. We can then write $y_t = F_\star z_t$ for a constant matrix $F_\star \in \mathbb{R}^{p \times \bar{n}}$, $\bar{n} := (L+1)m + Lp$. This is the well-known linear regression form of (4), and the z_t's are called the regression vectors [12, 13]. We can use this to parametrize the system behavior as

$$\mathcal{B}(F_\star) = \left\{ \begin{bmatrix} z \\ y \end{bmatrix} : \mathbb{Z}_+ \to \mathbb{R}^{\bar{n}+p} \,\middle|\, y_t = F_\star z_t,\, t \in \mathbb{Z}_+ \right\}. \tag{26}$$

Given observed data $\left(z_t^\mathrm{d}, y_t^\mathrm{d}\right)_{t=0}^{T-L-1}$, let $Y := [y_0^\mathrm{d}, \cdots, y_{T-L-1}^\mathrm{d}]$, and $H := [z_0^\mathrm{d}, \cdots z_{T-L-1}^\mathrm{d}]$. The following direction is easy to establish.

Lemma 7. *Any linear combination of columns of $\begin{bmatrix} H \\ Y \end{bmatrix}$ is a valid pair of regression and output vectors for the model* (4).

To prove the reverse direction, we make the connection to minimum-norm interpolation problems [14, 15, 16]: We have a vector-valued observation $Y = F_\star H$ of an unknown $F_\star$ and wish to approximate (or predict) $y = F_\star z$ for a given $z \in \mathbb{R}^{\bar{n}}$. We have the following error bound:

Lemma 8. *Let $\widehat{F}$ be the solution of the minimum-norm interpolation problem*

$$\begin{aligned} &\min \|F\|_\mathrm{F}^2 \\ &\textit{s.t. } FH = Y \end{aligned} \tag{27}$$

Then

$$\|y - \widehat{F}z\|_2 \leq J\|F_\star - \widehat{F}\|_\mathrm{F}, \tag{28}$$

where the quantity

$$J = \sup_{\substack{F \in \mathbb{R}^{p \times \bar{n}} \setminus \{0\} \\ FH=0}} \frac{\|Fz\|_2}{\|F\|_{\mathrm{F}}}. \tag{29}$$

is equal to zero if $z \in \mathrm{colspace}(H)$.

The scalar case ($p = 1$) was originally worked out in [14]. It was subsequently shown by Sard [16] that, for scalar observations, (28) holds with equality, and $J = \mathrm{dist}(z, \mathrm{colspace}(H))$ (this is a straightforward consequence of duality). The same result was obtained in the context of reproducing kernel Hilbert spaces by Liang and Recht [17]. The minimum-norm solution in (27) is given by $\widehat{F} = YH^\dagger$. The following is immediate:

Theorem 9. *Consider the input/output model* (4) *in regression form:* $y_t = F_\star z_t$ *for an unknown* $F_\star$*. Given* $T - L$ *data samples* $\left(z_t^{\mathrm{d}}, y_t^{\mathrm{d}}\right)_{t=0}^{T-L-1}$*, suppose that* H *has full row rank. If* (z, y) *is a valid pair of regression and output vectors, i.e.,* $y = F_\star z$*, then we can write* $y = Y\xi$ *for* $\xi = H^\dagger z$*.*

The requirement of H having full row rank is exactly the classical persistence of excitation condition for the regression vectors:

$$HH^\top = \sum_{t=0}^{T-L-1} z_t^{\mathrm{d}} (z_t^{\mathrm{d}})^\top \succ 0. \tag{30}$$

When the system (4) is completely reachable, the PE property of H is implied by the corresponding property of the inputs $(u_t^{\mathrm{d}})_{t=0}^{T-L-1}$ [12, 13]. Conceptually, the results of this section are in the spirit of the fundamental lemma, but phrased in terms of a different representation of the data by outputs and regression vectors that are consisting of inputs and past outputs. A related perspective on the fundamental lemma via regression in reproducing kernel Hilbert spaces was recently explored in [18].

VI. Conclusion

In this paper, we provided a unified behavioral perspective on subspace identification and minimum-norm interpolation approaches to simulation of linear systems and clarified the role of various structural assumptions on the system and on the data (both offline and online) in each setting. In future work, we plan to extend this behavioral approach to nonlinear systems.

Appendix I
Proofs

A. Proof of Lemma 2

Define matrices $\widetilde{A}_1, \widetilde{A}_2 \in \mathbb{R}^{Lp \times Lp}$ as

$$\begin{aligned}
\widetilde{A}_1 &= -\begin{bmatrix} A_L & A_{L-1} & A_{L-2} & \cdots & A_2 & A_1 \\ \mathbf{0} & A_L & A_{L-1} & \cdots & A_3 & A_2 \\ \mathbf{0} & \mathbf{0} & A_L & \cdots & A_4 & A_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & A_L \end{bmatrix}, \\
\widetilde{A}_2 &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ -A_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ -A_2 & -A_1 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -A_{L-1} & -A_{L-2} & \cdots & -A_1 & \mathbf{0} \end{bmatrix}.
\end{aligned} \tag{31}$$

Likewise, define matrices $\widetilde{B}_1, \widetilde{B}_2 \in \mathbb{R}^{L \times Lm}$ as

$$\begin{aligned}
\widetilde{B}_1 &= \begin{bmatrix} B_L & B_{L-1} & B_{L-2} & \cdots & B_2 & B_1 \\ \mathbf{0} & B_L & B_{L-1} & \cdots & B_3 & B_2 \\ \mathbf{0} & \mathbf{0} & B_L & \cdots & B_4 & B_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & B_L \end{bmatrix}, \\
\widetilde{B}_2 &= \begin{bmatrix} B_0 & \mathbf{0} & \cdots & \mathbf{0} \\ B_1 & \mathbf{0} & \cdots & \mathbf{0} \\ B_2 & B_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ B_{L-1} & B_{L-2} & \cdots & B_0 \end{bmatrix}.
\end{aligned} \tag{32}$$

Using (4), for $t = 0, 1, \ldots,$ we have

$$\begin{aligned}
\begin{bmatrix} y_{t+L} \\ \vdots \\ y_{t+2L-1} \end{bmatrix} =& \widetilde{A}_1 \begin{bmatrix} y_t \\ \vdots \\ y_{t+L-1} \end{bmatrix} + \widetilde{A}_2 \begin{bmatrix} y_{t+L} \\ \vdots \\ y_{t+2L-1} \end{bmatrix} \\
&+ \widetilde{B}_1 \begin{bmatrix} u_t \\ \vdots \\ u_{t+L-1} \end{bmatrix} + \widetilde{B}_2 \begin{bmatrix} u_{t+L} \\ \vdots \\ u_{t+2L-1} \end{bmatrix}.
\end{aligned} \tag{33}$$

Moving $y_{t+L:t+2L-1}$ to the left-hand side, we have

$$(I - \widetilde{A}_2) y_{t+L:t+2L-1} = \begin{bmatrix} \widetilde{A}_1 & \widetilde{B}_1 & \widetilde{B}_3 \end{bmatrix} \begin{bmatrix} u_{t:t+L-1} \\ y_{t:t+L-1} \\ u_{t+L:t+2L-1} \end{bmatrix} \tag{34}$$

Here, $I - \widetilde{A}_2$ is square, block lower triangular, and each diagonal block is the identity matrix I_p. Hence, it is invertible, and we obtain (7) with

$$K_\star = \begin{bmatrix} (I - \widetilde{A}_2)^{-1} \widetilde{B}_1 & (I - \widetilde{A}_2)^{-1} \widetilde{A}_1 & (I - \widetilde{A}_2)^{-1} \widetilde{B}_2 \end{bmatrix} \tag{35}$$

B. Proof of Thoerem 3

We start with the case when the identification is exact, i.e., $\widehat{K} = K_\star$. Using the expression for $K_\star$ in (35), we can write $\widehat{y}_{L:2L-1}$ in terms of $u_{0:L-1}$, $y_{0:L-1}$ and $u_{L:2L-1}$ as

$$\begin{aligned}\widehat{y}_{L:2L-1} =& K_\star \begin{bmatrix} u_{0:L-1} \\ y_{0:L-1} \\ u_{L:2L-1} \end{bmatrix} \\ =& (I - \widetilde{A}_2)^{-1}\widetilde{B}_1 \; u_{0:L-1} + (I - \widetilde{A}_2)^{-1}\widetilde{A}_1 \; y_{0:L-1} \\ &+ (I - \widetilde{A}_2)^{-1}\widetilde{B}_2 \; u_{L:2L-1}.\end{aligned} \tag{36}$$

Hence, we have

$$\begin{aligned}\widehat{y}_{L:2L-1} =& \widetilde{B}_1 \; u_{0:L-1} + \widetilde{A}_1 \; y_{0:L-1} \\ &+ \widetilde{B}_2 \; u_{L:2L-1} + \widetilde{A}_2 \; \widehat{y}_{L:2L-1}.\end{aligned} \tag{37}$$

Comparing with (33), we conclude that $\widehat{y}_{L:2L-1}$ is the output of the system (4) with input $u_{L:2L-1}$ and initial input-ouput pair $(u_{0:L-1}, y_{0:L-1})$.

For the general case, let $r := \begin{bmatrix} u_{0:L-1} \\ y_{0:L-1} \\ u_{L:2L-1} \end{bmatrix}$ and $R := \begin{bmatrix} U_\mathrm{p} \\ Y_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}$, where $r \in \mathrm{range}(R)$ by hypothesis. Since $K_\star$ is a particular solution of the least-squares problem, it can be written as

$$K_\star = \widehat{K} + M(I - RR^\dagger) \tag{38}$$

for some matrix M. By the properties of pseudoinverses, $RR^\dagger = RR^\intercal(RR^\intercal)^\dagger$. Consider the eigendecomposition of $RR^\intercal$ given by

$$RR^\intercal = U\Sigma U^\intercal = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^\intercal \\ U_2^\intercal \end{bmatrix} = U_1 \Sigma_1 U_1^\intercal. \tag{39}$$

Since $r \in \mathrm{range}(R)$, there exists some z such that $r = U_1 z$, and we have

$$\begin{aligned}(RR^\intercal)(RR^\intercal)^\dagger =& U_1\Sigma_1 U_1^\intercal \left(U_1 \Sigma_1^{-1} U_1^\intercal\right) r \\ =& U_1 I U_1^\intercal r \\ =& U_1 I U_1^\intercal (U_1 z) \\ =& U_1 z \\ =& r.\end{aligned} \tag{40}$$

Hence,

$$\widehat{K} r = K_\star r - M(I - (RR^\intercal)(RR^T)^\dagger) r = K_\star r. \tag{41}$$

Therefore, we have $\widehat{y}_{L:2L-1} = \widehat{K} r = K_\star r$. Proceeding as before, we conclude that $\widehat{y}_{L:2L-1}$ is still a valid output.

C. Proof of Theorem 4

By assumption, the input Hankel matrix $H_{2L+n}(u_{0:T-1})$ has full row rank. Consider the input Hankel matrix of depth $2L$ denoted by $H_{2L}(u_{0:T-1})$. Notice that $H_{2L}(u_{0:T-1})$ is the concatenation of U_p and U_f, i.e., $H_{2L}(u_{0:T-1}) = \begin{bmatrix} U_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix} =: U_{\mathrm{p,f}}$. Since $H_{2L}(u_{0:T-1})$ is a row submatrix of $H_{2L+n}(u_{0:T-1})$ and the latter has full row rank, $H_{2L}(u_{0:T-1})$ also has full row rank, i.e.,

$$\mathrm{rank}(H_{2L}(u_{0:T-1})) = 2Lm. \tag{42}$$

Hence, we have

$$\mathrm{rowspace}(U_\mathrm{p}) \cap \mathrm{rowspace}(U_\mathrm{f}) = \{0\}. \tag{43}$$

Since $\mathrm{rank}(\mathcal{C}_L) = n$, applying [19, Theorem 1], we have that $\mathrm{rank}\left(\begin{bmatrix} X_0 \\ U_{\mathrm{p,f}} \end{bmatrix}\right) = n + 2Lm$. Combined with (43), we have that the following two row submatrices also have full row rank:

$$\mathrm{rank}\left(\begin{bmatrix} X_0 \\ U_\mathrm{p} \end{bmatrix}\right) = Lm + n, \quad \mathrm{rank}\left(\begin{bmatrix} X_0 \\ U_\mathrm{f} \end{bmatrix}\right) = Lm + n. \tag{44}$$

Hence, we have

$$\mathrm{rowspace}(X_0) \cap \mathrm{rowspace}(U_\mathrm{p}) = \{0\}, \tag{45}$$

$$\mathrm{rowspace}(X_0) \cap \mathrm{rowspace}(U_\mathrm{f}) = \{0\}. \tag{46}$$

so Assumptions 1 and 2 in [10, Theorem 2, Chapter 2] hold.

D. Proof of Theorem 5

First, from [19, Theorem 1], the PE condition implies $\mathrm{rank}\left(\begin{bmatrix} X_0 \\ U_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}\right) = 2Lm + n.$

($\Rightarrow$): Suppose that $\begin{bmatrix} u_{0:2L-1} \\ y_{0:2L-1} \end{bmatrix}$ is a valid input-ouput trajectory of (20). Then there exists an initial state $\widehat{x}_0$ such that

$$y_{0:L-1} = \mathcal{O}_L \widehat{x}_0 + \mathcal{T}_L u_{0:L-1}. \tag{47}$$

Since $\mathrm{rank}\left(\begin{bmatrix} U_\mathrm{p} \\ U_\mathrm{f} \\ X_0 \end{bmatrix}\right) = 2Lm + n$, there exists some $\xi \in \mathbb{R}^{T-2L+1}$ such that

$$\begin{bmatrix} u_{0:L-1} \\ u_{L:2L-1} \\ \widehat{x}_0 \end{bmatrix} = \begin{bmatrix} U_\mathrm{p} \\ U_\mathrm{f} \\ X_0 \end{bmatrix} \xi. \tag{48}$$

Thus, we have

$$\begin{aligned}\begin{bmatrix} u_{0:L-1} \\ y_{0:L-1} \end{bmatrix} &= \begin{bmatrix} I & \mathbf{0} \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} u_{0:L-1} \\ \widehat{x}_0 \end{bmatrix} \\ &= \begin{bmatrix} I & \mathbf{0} \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} U_\mathrm{p} \\ X_0 \end{bmatrix} \xi \\ &= \begin{bmatrix} U_\mathrm{p} \\ Y_\mathrm{p} \end{bmatrix} \xi.\end{aligned} \tag{49}$$

Moreover, from the linear dynamics (20), we have

$$X_\mathrm{f} = A^L X_0 + \Delta_L U_\mathrm{p}. \tag{50}$$

Hence, at time L, we have

$$\begin{aligned}\widehat{x}_L =& A^L \widehat{x}_0 + \Delta_L u_{0:L-1} \\ =& \left(A^L X_0 + \Delta_L U_{\mathrm{p}}\right) \xi \\ =& X_{\mathrm{f}} \xi.\end{aligned} \tag{51}$$

Thus, starting from $\widehat{x}_L = X_{\mathrm{f}}\xi$ with input sequence $u_{L:2L-1} = U_{\mathrm{f}}\xi$, we have

$$\begin{aligned}\begin{bmatrix} u_{L:2L-1} \\ y_{L:2L-1} \end{bmatrix} &= \begin{bmatrix} I & \mathbf{0} \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} u_{L:2L-1} \\ \widehat{x}_L \end{bmatrix} \\ &= \begin{bmatrix} I & \mathbf{0} \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} U_{\mathrm{f}} \\ X_{\mathrm{f}} \end{bmatrix} \xi \\ &= \begin{bmatrix} U_{\mathrm{f}} \\ Y_{\mathrm{f}} \end{bmatrix} \xi.\end{aligned} \tag{52}$$

($\Leftarrow$) Since $\begin{bmatrix} u_{0:L-1} \\ y_{0:L-1} \end{bmatrix} = \begin{bmatrix} U_{\mathrm{p}} \\ Y_{\mathrm{p}} \end{bmatrix} \xi$, and $\begin{bmatrix} U_{\mathrm{p}} \\ Y_{\mathrm{p}} \end{bmatrix}$ is constrained by the matrix input-output relation (23), we have

$$\begin{bmatrix} u_{0:L-1} \\ y_{0:L-1} \end{bmatrix} = \begin{bmatrix} U_{\mathrm{p}} \\ Y_{\mathrm{p}} \end{bmatrix} \xi = \begin{bmatrix} I & \mathbf{0} \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} U_{\mathrm{p}} \\ X_0 \end{bmatrix} \xi. \tag{53}$$

Hence, we have

$$y_{0:L-1} = Y_{\mathrm{p}}\xi = \mathcal{O}_L X_0 \xi + \mathcal{T}_L U_{\mathrm{p}} \xi \tag{54}$$

Since $u_{0:L-1} = U_{\mathrm{p}}\xi$, we conclude that $y_{0:L-1}$ is the output from $\widehat{x}_0 := X_0\xi$ with input $u_{0:L-1}$. As $(u_{0:L-1}, y_{0:L-1})$ is a valid input-output trajectory, we can use the linear dynamics (20) and X_{f} defined in (50) to obtain the state vector at time L as

$$\widehat{x}_L = A^L \widehat{x}_0 + \Delta_L u_{0:L-1} = A^L X_0 \xi + \Delta_L U_{\mathrm{p}} \xi = X_{\mathrm{f}} \xi. \tag{55}$$

For the second segment, using the assumptions and the matrix input-output relation for U_{f}, Y_{f}, we have

$$\begin{bmatrix} u_{L:2L-1} \\ y_{L:2L-1} \end{bmatrix} = \begin{bmatrix} U_{\mathrm{f}} \\ Y_{\mathrm{f}} \end{bmatrix} \xi = \begin{bmatrix} I & \mathbf{0} \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} U_{\mathrm{f}} \\ X_{\mathrm{f}} \end{bmatrix} \xi. \tag{56}$$

Thus, we have

$$\begin{aligned}y_{L:2L-1} =& Y_{\mathrm{f}} \xi \\ =& \mathcal{O}_L X_{\mathrm{f}} \xi + \mathcal{T}_L U_{\mathrm{f}} \xi \\ =& \mathcal{O}_L \widehat{x}_L + \mathcal{T}_L u_{L:2L-1},\end{aligned} \tag{57}$$

where the last equality follows from (55). Therefore, we conclude that $y_{L:2L-1}$ is the output from $\widehat{x}_L = X_{\mathrm{f}}\xi$ with input $u_{L:2L-1}$.

E. Proof of Lemma 6

Recall that we have shown $\operatorname{rank}\left(\begin{bmatrix} X_0 \\ U_p \end{bmatrix}\right) = Lm + n$ in (44). From the state space representation, we have the matrix input-output relation

$$\begin{bmatrix} U_{\mathrm{p}} \\ Y_p \end{bmatrix} = \begin{bmatrix} I & \mathbf{0} \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix} \begin{bmatrix} U_{\mathrm{p}} \\ X_0 \end{bmatrix}. \tag{58}$$

Thus, under the assumption that $\mathcal{O}_L$ has full column rank, we have $\operatorname{rank}\left(\begin{bmatrix} U_{\mathrm{p}} \\ Y_{\mathrm{p}} \end{bmatrix}\right) = Lm + n$. On the other hand, formula (2.16) in [10] indicates that we have $\operatorname{rowspace}\left(\begin{bmatrix} U_{\mathrm{p}} \\ Y_{\mathrm{p}} \end{bmatrix}\right) \cap \operatorname{rowspace}(U_f) = \{0\}$. Since by the PE condition we also have $\operatorname{rank}(G^u_{\mathrm{ff}}) = \operatorname{rank}(U_{\mathrm{f}}) = Lm$. Hence, we have

$$\operatorname{rank}\left(\overline{G}_{\mathrm{pf}}\right) = \operatorname{rank}\left(\begin{bmatrix} U_{\mathrm{p}} \\ Y_{\mathrm{p}} \\ U_{\mathrm{f}} \end{bmatrix}\right) = 2Lm + n. \tag{59}$$

F. Proof of Lemma 7

Let $c = [c_0, \cdots, c_{T-L-1}]$, $\widetilde{z}_{-s} = \sum_{j=0}^{T-L-1} c_j z_{j-s}$, $\widetilde{y}_{-s} = \sum_{j=0}^{T-L-1} c_j y_{j-s}$, $\widetilde{u}_{-s} = \sum_{j=0}^{T-L-1} c_j u_{j-s}$ Since for each $j = 0, \cdots, T-L-1$, each column of $\begin{bmatrix} H \\ Y \end{bmatrix}$ is a valid trajectory of (4), we have

$$y_j + \sum_{k=1}^{L} A_k y_{j-k} = \sum_{l=0}^{L} B_l u_{j-l}. \tag{60}$$

Multiplying by c_j and summing over j, we have

$$\begin{aligned}&\sum_{j=0}^{T-L-1} c_j y_j + \sum_{k=1}^{L} A_k \left(\sum_{j=0}^{T-L-1} c_j y_{j-k}\right) \\ &= \sum_{l=0}^{L} B_l \left(\sum_{j=0}^{T-L-1} c_j u_{j-l}\right).\end{aligned} \tag{61}$$

Hence, we have

$$\widetilde{y}_0 + \sum_{k=1}^{L} A_k \widetilde{y}_{-k} = \sum_{l=0}^{L} B_l \widetilde{u}_{-l}. \tag{62}$$

That is, $(\widetilde{u}, \widetilde{y})$ satisfy equation (4).

G. Proof of Lemma 8

Equip $\mathcal{H} = \mathbb{R}^{p \times \bar{n}}$ with the Hilbert–Schmidt inner product

$$\langle F, G \rangle := \operatorname{tr}(F^{\top} G) \tag{63}$$

that induces the Frobenius norm. Given the linear measurement operator $M : \mathcal{H} \to (\mathbb{R}^p)^{T-L}$ by $MF := FH$, define the following subspaces of $\mathcal{H}$:

$$\mathcal{N} := \ker(M), \tag{64}$$

$$\mathcal{M} := \mathcal{N}^{\perp} = \{F \in \mathcal{H} : \langle F, G \rangle = 0,\, G \in \mathcal{N}\}. \tag{65}$$

The adjoint operator $M^* : (\mathbb{R}^p)^{T-L} \to \mathcal{H}$ is given by

$$M^*(y_0, \ldots, y_{T-L-1}) = \sum_{t=0}^{T-L-1} y_t z_t^{\top}, \qquad y_t \in \mathbb{R}^p. \tag{66}$$

Since all spaces here are finite-dimensional, the range $\operatorname{range}(M)$ is closed, and therefore $\mathcal{M} = \mathcal{N}^{\perp} = \operatorname{range}(M^*)$ [20, Section 6.6, Theorem 2].

Let $\Pi_{\mathcal{M}}$ denote the orthogonal projection onto $\mathcal{M}$. Then $\widehat{F} := \Pi_{\mathcal{M}} F_{\star}$ is the unique solution of the minimum-norm interpolation problem (27), and there exists a linear operator E, such that $\widehat{F} = E F_{\star} H = EY$ [16, Theorems 1–4].

Moreover, for any linear operator $T : \mathcal{H} \to \mathbb{R}^p$ we have the error estimate

$$\|T\widehat{F} - TF_\star\|_2 \leq \|T \restriction N\| \cdot \|F_\star - F\|_{\mathrm{F}}, \tag{67}$$

where

$$\|T \restriction \mathcal{N}\| = \sup_{\substack{F\in\mathcal{N} \\ \|F\|_{\mathrm{F}}=1}} \|TF\|_2 \tag{68}$$

is the operator norm of the restriction of T to $\mathcal{N}$ [16, Lemma 4]. We specialize this to the operator $T_z F := Fz$:

$$\|T_z \restriction \mathcal{N}\| = \sup_{\substack{F\in\mathcal{N} \\ \|F\|_{\mathrm{F}}=1}} \|Fz\|, \tag{69}$$

and the expression for J in (29) is immediate from this. By the characterization of $\mathcal{M} = \mathrm{range}(M^*)$, $\|T_z \restriction \mathcal{N}\| = 0$ whenever $z \in \mathrm{colspace}(H)$.

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 deepc," 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] W. Favoreel, B. De Moor, and M. Gevers, "SPC: subspace predictive control," *IFAC Proceedings Volumes*, vol. 32, no. 2, pp. 4004–4009, 1999.

[10] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: Theory—Implementation—Applications*. Springer Science and Business Media, 2012.

[11] R. J. P. de Figueiredo, A. Caprihan, and A. N. Netravali, "On optimal modeling of systems," *Journal of Optimization Theory and Applications*, vol. 11, no. 1, pp. 68–83, 1973.

[12] J. Moore, "Persistence of excitation in extended least squares," *IEEE Transactions on Automatic Control*, vol. 28, no. 1, pp. 60–68, 1983.

[13] M. Green and J. B. Moore, "Persistence of excitation in linear systems," *Systems & Control Letters*, vol. 7, no. 5, pp. 351–360, 1986.

[14] M. Golomb and H. F. Weinberger, "Optimal approximation and error bounds," in *On Numerical Approimxation*, R. E. Langer, Ed. Madison, WI: University of Wisconsin Press, 1959, pp. 117–190.

[15] A. Sard, "Optimal approximation," *Journal of Functional Analysis*, vol. 1, no. 2, pp. 222–244, 1967.

[16] ——, "Approximation based on nonscalar observations," *Journal of Approximation Theory*, vol. 8, no. 4, pp. 315–334, 1973.

[17] T. Liang and B. Recht, "Interpolating classifiers make few mistakes," *Journal of Machine Learning Research*, vol. 24, no. 20, pp. 1–27, 2023.

[18] 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.

[19] 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.

[20] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: John Wiley & Sons, 1969.