2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

Data-Driven State Estimation for Linear Systems

Vikas Kumar Mishra, Sandesh Athni Hiremath, and Naim Bajcinca

***Abstract*— We study the problem of estimating the states of a linear system based on measured data. We investigate the problem in both deterministic and stochastic settings. In the deterministic case, we develop data-driven conditions under which we can reconstruct state trajectories uniquely. Also, we discuss the case in which we have some missing data in the given input/output measurements. In the stochastic case, we develop a Kalman filter-like algorithm to recursively estimate both states and outputs. Finally, we consider a multi-input multi-output system to elucidate the developed results.**

***Index Terms*— State reconstruction/estimation, data-driven approach, missing data, behavioral system theory, LTI systems.**

I. Introduction

State estimation is a classical problem in systems and control. Luenberger observer [1] (in the deterministic setting) or Kalman filter [2] (in the stochastic setting) is the state-of-the-art for reconstructing or estimating the states of a dynamic system. Traditionally, both Luenberger observers and Kalman filters require model knowledge for estimating the states. However, due to the increasing complexity of modern control systems, it is not always possible to derive a model using first principles or system identification algorithms. With that in mind, this paper considers the problem of state estimation directly from measured data, without explicitly identifying a state-space model. Our approach to tackling this problem is inspired by the behavioral system theory of Willems [3]. In particular, it relies on a seminal result that is known as the *fundamental lemma* [4, Theorem 1]. According to this result, the subspace of all possible input/output trajectories of a controllable linear time-invariant (LTI) system can be obtained by a single sufficiently rich input/output trajectory.

As the fundamental lemma provides sufficient conditions under which a completely data-driven representation can replace the parametric state-space model, this result has become instrumental in the development of recent data-driven control methods. Necessary and sufficient conditions for the fundamental lemma have been stated and proven in [5], see also [6] for further discussion. Furthermore, it has been extended to uncontrollable LTI systems [7], [8] and to the case in which multiple trajectories of the system are given [9]. It has also been extended to classes of nonlinear systems such as Volterra systems [10], NARX systems [11], and Wiener and Hammerstein systems [12]. For an overview, we refer to [13]. The fundamental lemma has directly or indirectly been used in solving several control problems: data-driven simulation or prediction [14], data-driven control in the behavioral setting [15], [16], data-driven stabilizing controllers [17], data-driven predictive control [18], [19], [20], [21], [22], data-driven input reconstruction [23], [24] and moving horizon estimation [25]. For further applications, specifically, in power systems, we refer the readers to [26].

Furthermore, a Kalman filter-like algorithm, based on the fundamental lemma, to estimate only the outputs is discussed in [27, Section IV], whereas an extended Kalman filter-like algorithm has been proposed to estimate the fictitious state (sequence of past outputs) for the purposes of data-driven predictive control in [28, Section IV]. Other contributions that consider the estimation of states include [29], [30]. Note that the approach in [29] is based on the fundamental lemma, whereas the approach in [30] relies on the data informativity framework proposed in [31]. In this paper, we are interested in estimating the states of a linear system in deterministic as well as stochastic settings. Note that we use the terminology reconstruction in the deterministic case and estimation in the stochastic case. The main contributions of this paper are:

1) we state and prove an extension of the fundamental lemma (see Theorem 3), where state trajectories are also considered, and using this result we develop Algorithm 1 that reconstructs the state trajectory from a given input/output trajectory;
2) we extend Algorithm 1 to the case, where we have missing data in input/output trajectories (see Algorithm 2);
3) we extend Theorem 3 in stochastic setting (see Theorem 8) and use it to develop Kalman filter-like recursive algorithm for estimating both outputs and states based on measured data of a noisy system (see Algorithm 3).

The rest of this paper is organized as follows. We provide our notation and necessary preliminaries in Section II. Section III presents the problem considered in this paper. Section IV deals with the state reconstruction in the noise-free setting. Section V deals with the state reconstruction in case of missing data. Section VI develops a data-driven Kalman filter-like algorithm for both output and state estimation. Section VII illustrates our results on a multi-input multi-output (MIMO) system. Finally, Section VIII offers concluding remarks and future research directions.

II. Notation and Preliminaries

The set of real $k \times m$ matrices is denoted by $\mathbb{R}^{k\times m}$. The transpose and the Moore-Penrose pseudo-inverse of $A \in$

This work was supported by AiF within the scope of the project VirMan with the grant number KK5048411LF1 and by DFG priority program (SPP 2364) - Autonomous processes in particle technology - GZ: BA 5511/9-1.

The authors are with the Department of Mechanical and Process Engineering, RPTU Kaiserslautern, Gottlieb-Daimler-Straße 42, 67663 Kaiserslautern, Germany (email: {vikas.mishra,sandesh.hiremath naim.bajcinca}@rptu.de).

$\mathbb{R}^{k\times m}$ are denoted by $A^\top$ and $A^\dagger$, respectively. The identity matrix of dimension k/appropriate is denoted by I_k/I and the zero matrix of appropriate dimensions is denoted by 0. We define

$$\text{col}(A_1, A_2, \dots, A_r) := \begin{bmatrix} A_1^\top & A_2^\top & \dots & A_r^\top \end{bmatrix}^\top$$

provided the matrices have the same number of columns. With z_d we denote a generic offline trajectory in discrete-time (*data*) of length $T \in \mathbb{N}$ defined as[1]

$$z_d := \big(z_d(t), z_d(t+1), \dots, z_d(t+T-1)\big) \in (\mathbb{R}^q)^T.$$

Associated with z_d, we define a shifted trajectory as

$$z_d^+ := \big(z_d(t+1), z_d(t+2), \dots, z_d(t+T)\big) \in (\mathbb{R}^q)^T.$$

Further, we denote a generic trajectory of length $L \in \mathbb{N}$ as $z|_{[t,t+L-1]} := \text{col}(z(t), z(t+1), \dots, z(t+L-1))$. For the sake of simplicity, we denote $f|_L := f|_{[1,L]}$ and $f|_{L_t} = f_{t-L_p,t+L_f}$ for $L = L_p + L_f + 1, L_p, L_f \in \mathbb{N}$. We denote by $z_{s|t}$ the predicted value of a variable z at time $s \geq t$ given the information at time t. We denote any symmetric positive definite matrix A by $A > 0$. $\mathbb{S}^q := \{A \in \mathbb{R}^{q\times q} : A > 0\}$ denotes the set of symmetric positive definite matrices of size $q \times q$. The abbreviation rv(s) stands for random vector(s). The triplet $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a complete probability space with Ω being the sample space, $\mathcal{F}$ the complete sigma algebra, $\mathbb{P}$ a probability measure. The symbol $\mathbb{E}$ denotes the expectation operator with respect to the measure $\mathbb{P}$. For an rv $w \in \mathbb{R}^q$, we denote its expected value $\mathbb{E}[w]$ as $\bar{w}$. For any centered rvs $w \in \mathbb{R}^q$ and $v \in \mathbb{R}^p$, we denote $\text{cov}(w,v) = \mathbb{E}\big[wv^\top\big] \in \mathbb{R}^{q\times p}$ and $\text{var}(w) = \text{cov}(w,w) \in \mathbb{R}^{q\times q}$ the covariance matrices.

Now, we recall the notion of persistency of excitation [4].

Definition 1 (Persistency of excitation): A q-variate time series z_d is persistently exciting of order $L \in \mathbb{N}$ if the Hankel matrix with L-block rows defined as

$$\mathcal{H}_L(z_d) := \begin{bmatrix} z_d(t) & z_d(t+1) & \cdots & z_d(t+T-L) \\ z_d(t+1) & z_d(t+2) & \cdots & z_d(t+T-L+1) \\ \vdots & \vdots & \ddots & \vdots \\ z_d(t+L-1) & z_d(t+L) & \cdots & z_d(t+T-1) \end{bmatrix}$$

has full row rank, i.e., its rank is qL.

Finally, $\mathcal{H}|_{[l_1,l_2]}(z_d)$ represents a submatrix of $\mathcal{H}_L(z_d)$ starting from the block row l_1 to block row l_2 included.

III. Problem Statement

Consider the following minimal discrete-time linear time-invariant system

$$x_{t+1} = Ax_t + Bu_t \tag{1a}$$
$$y_t = Cx_t + Du_t. \tag{1b}$$

Here, at each time instant $t \in \mathbb{N}$, $x_t \in \mathbb{R}^n$ is the state vector, $u_t \in \mathbb{R}^m$ is the control input vector, and $y_t \in \mathbb{R}^p$ is the output vector. Matrices A, B, C, D are of appropriate dimensions. We assume that the matrices A, B, C, D are unknown; however, we have access to input/output or input/state/output data that are generated by the system. The aim is to reconstruct the state of the system based on measured data. The noisy case is considered in Section VI.

[1] Trajectories x_d, u_d, and y_d will be defined similarly.

IV. Data-driven representation of an LTI system

We begin with by recalling the fundamental lemma [4, Theorem 1].

Lemma 2 (Fundamental lemma): Assume that system (1) is controllable and, given $L \in \mathbb{N}$ with $L > n$, the observed trajectory $\text{col}(u_d, y_d)$ is such that u_d is persistently exciting of order $L+n$. Then, $\text{col}(u|_{[t,t+L-1]}, y|_{[t,t+L-1]})$ is a trajectory of system (1) if and only if there exists $g_t \in \mathbb{R}^{T-L+1}$ such that

$$\begin{bmatrix} \mathcal{H}_L(u_d) \\ \mathcal{H}_L(y_d) \end{bmatrix} g_t = \begin{bmatrix} u|_{[t,t+L-1]} \\ y|_{[t,t+L-1]} \end{bmatrix}.$$

We now show that the fundamental lemma due to Willems et al. can be straightforwardly extended to include the state trajectories as well. Note that by L-long input/state/output trajectory of system (1) we mean $\text{col}(u|_{[t,t+L-1]}, x|_{[t+1,t+L]}, y|_{[t,t+L-1]})$. Note the indices on different variables.

Theorem 3: Let the system (1) be controllable. Let u_d be persistently exciting of order $n+L$. Then, $\text{col}(u|_{[t,t+L-1]}, x|_{[t+1,t+L]}, y|_{[t,t+L-1]})$ is a trajectory of (1) if and only if there exists $g_t \in \mathbb{R}^{T-L+1}$ such that

$$\begin{bmatrix} \mathcal{H}_L(u_d) \\ \mathcal{H}_L(x_d^+) \\ \mathcal{H}_L(y_d) \end{bmatrix} g_t = \begin{bmatrix} u|_{[t,t+L-1]} \\ x|_{[t+1,t+L]} \\ y|_{[t,t+L-1]} \end{bmatrix}. \tag{2}$$

Proof: 'If' part: Let (2) hold. Then, by superposition, $\text{col}(u|_{[t,t+L-1]}, x|_{[t+1,t+L]}, y|_{[t,t+L-1]})$ is a trajectory of (1).

'Only if' part: Because the system (1) is controllable and u_d is persistently exciting of order $n+L$, the matrix

$$\begin{bmatrix} \mathcal{H}_L(u_d) \\ \mathcal{H}|_{[1,1]}(x_d) \end{bmatrix} \tag{3}$$

has full row rank [9, Theorem 1]. Thus, there exists a solution $g_t \in \mathbb{R}^{T-L+1}$ to the following system of linear equations

$$\begin{bmatrix} \mathcal{H}_L(u_d) \\ \mathcal{H}|_{[1,1]}(x_d) \end{bmatrix} g_t = \begin{bmatrix} u|_{[t,t+L-1]} \\ x_t \end{bmatrix}. \tag{4}$$

From system (1), we have

$$\begin{bmatrix} u|_{[t,t+L-1]} \\ x|_{[t+1,t+L]} \\ y|_{[t,t+L-1]} \end{bmatrix} = \underbrace{\begin{bmatrix} I_{mL} & 0 \\ \mathcal{N}_L & \mathcal{P}_L \\ \mathcal{T}_L & \mathcal{O}_L \end{bmatrix}}_{=:\mathbf{M}_L} \begin{bmatrix} u|_{[t,t+L-1]} \\ x_t \end{bmatrix}, \tag{5}$$

where

$$\mathcal{N}_L = \begin{bmatrix} B & 0 & \cdots & 0 \\ AB & B & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A^{L-1}B & A^{L-2}B & \cdots & B \end{bmatrix}, \; \mathcal{O}_L = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{L-1} \end{bmatrix},$$

$$\mathcal{P}_L = \begin{bmatrix} A \\ A^2 \\ \vdots \\ A^L \end{bmatrix}, \; \mathcal{T}_L = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{L-2}B & CA^{L-3}B & \cdots & D \end{bmatrix}. \tag{6}$$

From equations (4) and (5), we have

$$\begin{bmatrix} u|_{[t,t+L-1]} \\ x|_{[t+1,t+L]} \\ y|_{[t,t+L-1]} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_L(u_d) \\ \mathcal{H}_L(x_d^+) \\ \mathcal{H}_L(y_d) \end{bmatrix} g_t. \tag{7}$$

This completes the proof. ∎

Based on the above theorem, we develop the following algorithm for state reconstruction.

Algorithm 1: Data-driven state reconstruction.

Input: Data $\operatorname{col}(u_d, x_d^+, y_d)$ and measurements $\operatorname{col}(u|_{[t,t+L-1]}, y|_{[t,t+L-1]})$.

Output: State trajectory $x|_{[t+1,t+L]}$.

1: Compute a g from

$$\begin{bmatrix} \mathcal{H}_L(u_d) \\ \mathcal{H}_L(y_d) \end{bmatrix} g = \begin{bmatrix} u|_{[t,t+L-1]} \\ y|_{[t,t+L-1]} \end{bmatrix}. \tag{8}$$

2: Compute

$$x|_{[t+1,t+L]} = \mathcal{H}_L(x_d^+)g. \tag{9}$$

Now, we assume that

(C0) $\ker(\operatorname{col}(\mathcal{H}_L(u_d), \mathcal{H}_L(y_d))) \subseteq \ker(\mathcal{H}_L(x_d^+))$.

Theorem 4: Let the system (1) be controllable, u_d be persistently exciting of order $n+L$, and condition (C0) hold. Then, Algorithm 1 gives a unique state trajectory.

Proof: The proof is straightforward and is skipped. ∎

Remark 1: Evidently, Algorithm 1 relies on state data. However, these data x_d, x_d^+ are collected offline and by using Algorithm 1, we can determine state trajectories at any point in time or in online experiments. Note that related prior contributions [29], [30] require state data as well.

V. Missing Trajectories

In this section, we consider the problem of reconstructing the state trajectory in case the measured input/output trajectory $\operatorname{col}(u|_L, y|_L)$ has some missing entries or outliers. Note that we treat the outliers as missing entries. Also, note that, for the simplicity of exposition, we consider $\operatorname{col}(u|_L, y|_L)$ instead of $\operatorname{col}(u|_{[t,t+L-1]}, y|_{[t,t+L-1]})$. We denote the known trajectory by $(u|_{K_1}, y|_{K_2})$, where $K_1 \subseteq \{1, 2, \ldots, mL\}$ and $K_2 \subseteq \{1, 2, \ldots, pL\}$. Consequently, we denote the missing trajectory by $(u|_{M_1}, y|_{M_2})$, where $M_1 = \{1, 2, \ldots, mL\} - K_1$ and $M_2 = \{1, 2, \ldots, pL\} - K_2$. Correspondingly, we denote by $\mathcal{H}_{K_1}(u_d)$ and $\mathcal{H}_{K_2}(y_d)$ as the submatrices of $\mathcal{H}_L(u_d)$ and $\mathcal{H}_L(y_d)$, respectively, obtained by deleting the rows that correspond to the sets M_1 and M_2.

To tackle the above problem, we first reconstruct the full-length input/output trajectory $(u|_L, y|_L)$, based on the known trajectory $(u|_{K_1}, y|_{K_2})$. In other words, we find out the missing entries $(u|_{M_1}, y|_{M_2})$ of the given trajectory. Then, we use our Algorithm 1 to reconstruct the state trajectory, based on this reconstructed full-length input/output trajectory.

Computing the missing entries of a given trajectory has been more recently studied in [32], where this problem has been referred to as "data-driven dynamic interpolation." In a data-driven dynamic interpolation problem, we compute the missing entries of a partially known trajectory. Thus, we first complete the given partially known input/output trajectory using the data-driven dynamic interpolation algorithm [32, Algorithm 1] and then leverage Algorithm 1 developed in the previous section to reconstruct the states of the underlying system. Combining them, we develop the following algorithm to reconstruct the states of an LTI system from partially known input/output measurements.

Algorithm 2: Data-driven state reconstruction from partially known measurements.

Input: Data $\operatorname{col}(u_d, x_d^+, y_d)$ and measurements $\operatorname{col}(u|_L, y|_L)$ with missing entries.

Output: Complete input/output trajectory $\operatorname{col}(\hat{u}|_L, \hat{y}|_L)$ and state trajectory $x|_{[2,L+1]}$.

1: Compute a g_m from

$$\begin{bmatrix} \mathcal{H}_{K_1}(u_d) \\ \mathcal{H}_{K_2}(y_d) \end{bmatrix} g_m = \begin{bmatrix} u|_{K_1} \\ y|_{K_2} \end{bmatrix}. \tag{10}$$

2: Compute

$$\begin{bmatrix} \hat{u}|_L \\ \hat{y}|_L \end{bmatrix} = \begin{bmatrix} \mathcal{H}_L(u_d) \\ \mathcal{H}_L(y_d) \end{bmatrix} g_m. \tag{11}$$

3: Compute a g from

$$\begin{bmatrix} \mathcal{H}_L(u_d) \\ \mathcal{H}_L(y_d) \end{bmatrix} g = \begin{bmatrix} \hat{u}|_L \\ \hat{y}|_L \end{bmatrix}. \tag{12}$$

4: Compute

$$x|_{[2,L+1]} = \mathcal{H}_L(x_d^+)g. \tag{13}$$

We remark that the data-driven dynamic interpolation problem discussed above (computing the missing entries—Steps 1 & 2 of Algorithm 2) need not be always solvable. Necessary and sufficient conditions for the unique solvability of the problem are given by (*cf.* [32, Section 3]):

(C1) $\operatorname{rank} \mathcal{H}_L(\operatorname{col}(u_d, y_d)) = n + mL$, for $L > n$;

(C2) $\operatorname{rank} \operatorname{col}(\mathcal{H}_{K_1}(u_d), \mathcal{H}_{K_2}(y_d)) = \operatorname{rank} \begin{bmatrix} \operatorname{col}(\mathcal{H}_{K_1}(u_d), \mathcal{H}_{K_2}(y_d)) & \operatorname{col}(u|_{K_1}, y|_{K_2}) \end{bmatrix}$;

(C3) $\operatorname{rank} \operatorname{col}(\mathcal{H}_{K_1}(u_d), \mathcal{H}_{K_2}(y_d)) = \operatorname{rank} \mathcal{H}_L(\operatorname{col}(u_d, y_d)) = n + mL$.

Note that (C1) is a standard assumption, which is equivalent to the identifiability of the system and can be enforced by an input sequence that is persistently exciting of order $n + L$ [21, Theorem 2]. Condition (C2) corresponds to the consistency condition for a solution to the system of equations (10). Finally, (C3) provides a unique solution to the data-driven dynamic interpolation problem. Thus, we have the following result.

Lemma 5: Let conditions (C0)–(C3) hold. Then, Algorithm 2 gives a unique state trajectory.

A. Missing data in offline trajectories

We now address the problem of reconstructing the state variables when we have missing data or outliers in offline trajectories. Under the assumption that the data are generated by an LTI system, a long trajectory with missing entries or outliers (which we treat as missing entries) can be considered as a collection of several short trajectories. Thus, we tackle this problem by leveraging the fundamental lemma for multiple trajectories [9]. Now, we recall the notion of a *collectively persistently exciting* set of signals of order L.

Definition 6: [9] Consider the time series z_{d_i} of length T_i for $i = 1, 2, \ldots, \alpha$, where α is the number of different time series. We say that the set of time series z_{d_i} of length T_i for $i = 1, 2, \ldots, \alpha$ is *collectively persistently exciting* of order $L \leq T_i$ if the mosaic Hankel matrix

$$\begin{bmatrix}\mathcal{H}_L(z_{d_1}) & \mathcal{H}_L(z_{d_2}) & \cdots & \mathcal{H}_L(z_{d_\alpha})\end{bmatrix}$$

is of full row rank.

Theorem 7: Let the system (1) be controllable. Let u_{d_i} of length T_i for $i = 1, 2, \ldots, \alpha$, where α is the number of different time series, be collectively persistently exciting of order $n + L$. Then, $\mathrm{col}(u|_{[t,t+L-1]}, x|_{[t+1,t+L]}, y|_{[t,t+L-1]})$ is a trajectory of (1) if and only if there exists g_t such that

$$\begin{bmatrix}\mathcal{H}_L(u_{d_1}) & \cdots & \mathcal{H}_L(u_{d_\alpha})\\ \mathcal{H}_L(x^+_{d_1}) & \cdots & \mathcal{H}_L(x^+_{d_\alpha})\\ \mathcal{H}_L(y_{d_1}) & \cdots & \mathcal{H}_L(y_{d_\alpha})\end{bmatrix} g_t = \begin{bmatrix}u|_{[t,t+L-1]}\\ x|_{[t+1,t+L]}\\ y|_{[t,t+L-1]}\end{bmatrix}. \quad (14)$$

Remark 2: Of course, if the inputs are collectively persistently exciting, then one can first identify a model and then use that model to recover the missing offline data, *cf.* [9, Section IV.A]. Furthermore, one can also use that model to reconstruct the states. However, by exploiting Theorem 7, in view of Algorithm 1, one can directly reconstruct the states without explicitly identifying a state-space model or recovering the missing offline data.

Remark 3: If we have missing entries in both offline data and online measured trajectories, we can again reconstruct the states by using Theorem 7 and Algorithm 2.

VI. Kalman filter-type recursive estimation of state and output trajectories

For each $t \in \mathbb{N}$ fixed, let μ_t, ν_t be rvs defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Consider the following linear stochastic system

$$x_{t+1} = Ax_t + Bu_t + \mu_t \quad (15a)$$

$$y_t = Cx_t + Du_t + \nu_t \quad (15b)$$

with $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^m$ and $y_t \in \mathbb{R}^p$ for $t \in \mathbb{N}$. Here we assume that initial state x_0, process noise μ_t, and the measurement noise ν_t are Gaussian distributed with the latter two having mean zero. Particularly, we have that $x_0 = \bar{x}_0 + \mu_0$ with $\mu_0 \sim \mathcal{N}(0, \Sigma_0)$, $\mu_t \sim \mathcal{N}(0, \Sigma_{\mu_t})$, and $\nu_t \sim \mathcal{N}(0, \Sigma_{\nu_t})$ for $\Sigma_0 \in \mathbb{S}^n$, $\Sigma_{\nu_t} \in \mathbb{S}^p$. Additionally, $u_t = \bar{u}_t + \eta_t$ where $\eta_t \sim \mathcal{N}(0, \Sigma_{\eta_t})$, with $\Sigma_{\eta_t} \in \mathbb{S}^m$, is a rv independent of μ_t and ν_t. Thus, the L-long trajectories, with $L = L_p + L_f + 1$, $L_p, L_f \in \mathbb{N}$, of the system (15) can be written as follows:

$$\begin{bmatrix}u|_{[t-L_p,t+L_f]}\\ x|_{[t-L_p+1,t+L_f+1]}\\ y|_{[t-L_p,t+L_f]}\end{bmatrix} = \underbrace{\begin{bmatrix}I_{mL} & 0\\ \mathcal{N}_L & \mathcal{P}_L\\ \mathcal{T}_L & \mathcal{O}_L\end{bmatrix}}_{=:\mathbf{M}_L} \begin{bmatrix}\bar{u}|_{[t-L_p,t+L_f]}\\ \bar{x}_{t-L_p}\end{bmatrix} + \underbrace{\begin{bmatrix}I_{mL} & 0 & 0\\ \mathcal{N}_L & I_{nL} & 0\\ \mathcal{D}_L & \mathcal{S}_L & I_{pL}\end{bmatrix}}_{=:\mathcal{K}_L} \begin{bmatrix}\eta|_{[t-L_p,t+L_f]}\\ \mu|_{[t-L_p,t+L_f]}\\ \nu|_{[t-L_p,t+L_f]}\end{bmatrix}, \quad (16)$$

where $\mathcal{N}_L$, $\mathcal{O}_L$, $\mathcal{P}_L$, $\mathcal{T}_L$ are as defined in (6) and

$$\mathcal{S}_L = \begin{bmatrix}C & 0 & \cdots & 0\\ CA & C & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ CA^{L-1} & CA^{L-2} & \cdots & C\end{bmatrix}, \quad \mathcal{D}_L = \begin{bmatrix}D & & \\ & \ddots & \\ & & D\end{bmatrix}.$$

Moreover, we have that

$$\eta|_{L_t} := \eta|_{[t-L_p,t+L_f]} \sim \mathcal{N}(0, \Sigma_{\eta|_{L_t}}),$$

$$\mu|_{L_t} := \mu|_{[t-L_p,t+L_f]} \sim \mathcal{N}(0, \Sigma_{\mu|_{L_t}}),$$

$$\nu|_{L_t} := \nu|_{[t-L_p,t+L_f]} \sim \mathcal{N}(0, \Sigma_{\nu|_{L_t}}),$$

where $\Sigma_{\mu|_{L_t}} \in \mathbb{S}^{nL}$, $\Sigma_{\nu|_{L_t}} \in \mathbb{S}^{pL}$, and $\Sigma_{\eta|_{L_t}} \in \mathbb{S}^{mL}$. Since μ_t, ν_t and η_t are from the same probability space, (16) can be written as

$$\begin{aligned} z|_{L_t} &= \mathbf{M}_L \begin{bmatrix}\bar{u}|_{L_t}\\ \bar{x}_{t-L_p}\end{bmatrix} + \mathcal{K}_L \tilde{\kappa}|_{L_t}\\ &= \mathbf{M}_L \begin{bmatrix}\bar{u}|_{L_t}\\ \bar{x}_{t-L_p}\end{bmatrix} + \kappa|_{L_t}. \end{aligned} \quad (17)$$

Here,

$$z|_{L_t} := z|_{[t-L_p,t+L_f]} := \begin{bmatrix}u|_{L_t}\\ x|_{L_t}\\ y|_{L_t}\end{bmatrix},$$

$$\kappa|_{L_t} := \begin{bmatrix}\eta|_{L_t}\\ \mu|_{L_t}\\ \theta|_{L_t}\end{bmatrix} = \mathcal{K}_L \begin{bmatrix}\eta|_{L_t}\\ \mu|_{L_t}\\ \nu|_{L_t}\end{bmatrix} = \mathcal{K}_L \tilde{\kappa}|_{L_t}, \quad (18)$$

with

$$\theta|_{L_t} := \mathcal{D}_L \eta|_{L_t} + \mathcal{S}_L \mu|_{L_t} + \nu|_{L_t},$$

which is again a Gaussian rv with

$$\theta|_{L_t} \sim \mathcal{N}(0, \Sigma_{\theta|_{L_t}}),$$

and

$$\Sigma_{\theta|_{L_t}} := \mathcal{D}_L \Sigma_{\eta|_{L_t}} \mathcal{D}_L^\top + \mathcal{S}_L \Sigma_{\mu|_{L_t}} \mathcal{S}_L^\top + \Sigma_{\nu|_{L_t}}.$$

With this, we denote the noise-parameterized behavior of the system as

$$\mathcal{B}_{L_t}(\kappa|_{L_t}) := \Big\{ z|_{L_t} : z|_{L_t} = \mathbf{M}_L \begin{bmatrix}\bar{u}|_{L_t}\\ \bar{x}_{t-L_p}\end{bmatrix} + \kappa|_{L_t} \text{ for } \bar{x}_{t-L_p} \in \mathbb{R}^n \Big\}.$$

Theorem 8: Consider system (15). Let u_d be persistently exciting of order $n + L$. Then $z|_{L_t} \in \mathcal{B}_{L_t}(\kappa|_{L_t})$ if and only if there exists $\bar{g}_t \in \mathbb{R}^{T-L+1}$ and $\kappa|_{L_t} \sim \mathcal{N}(0, \Sigma_{\kappa_{L_t}})$ such that

$$z|_{L_t} = \mathcal{H}_L(z_d)\bar{g}_t + \kappa|_{L_t}. \quad (19)$$

Proof: The proof follows similarly to the lines of Theorem 3 from [21] and Theorem 3 above. The interesting part that is relevant to the latter content is the following. For $\bar{g}_t$ such that $\bar{z}_t = \mathcal{H}_L(z_d)\bar{g}_t$, we can obtain a noisy trajectory by considering a noisy perturbed $g_t = \bar{g}_t + w_t$ for $w_t \sim \mathcal{N}(0, \Sigma_{w_t})$ such that

$$\begin{aligned} \bar{z}|_{L_t} + \kappa|_{L_t} &= \mathcal{H}_L(z_d)[\bar{g}_t + w_t] \\ &= \mathcal{H}_L(z_d)\bar{g}_t + \mathcal{H}_L(z_d)w_t \\ \Rightarrow z|_{L_t} &= \mathcal{H}_L(z_d)\bar{g}_t + \mathcal{H}_L(z_d)w_t \end{aligned} \tag{20}$$

with $\kappa|_{L_t} = \mathcal{H}_L(z_d)w_t$ and $\Sigma_{\kappa|_{L_t}} = \mathcal{H}_L(z_d)\, \Sigma_{w_t}\, \mathcal{H}_L^\top(z_d)$. ■

A. Recursive data-driven estimation

In this section, we shall present a recursive method for estimating the state and output of the system (15). To this end, we rely on the stochastic fundamental lemma (cf. Theorem 8) to first obtain a recursive relation for $(g_t)_{t\geq 0}$.

Theorem 9: Let $\bar{g}_t \in \mathbb{R}^{T-L+1}$ be a time dependent vector and $(w_t)_{t\geq 0}$ be a stochastic process taking values in $\mathbb{R}^{T-L+1}$ such that $\bar{g}_t$ and w_t satisfies (19) in the sense of (20). If

$$w_t = w_{t-1} + \delta_t \tag{21}$$

with $(\delta_t)_{t\geq 0}$ denoting independent increments of w_t and

$$\begin{aligned} \bar{x}|_{[t-L_p+1,t]} &= \mathcal{H}|_{[1,L_p]}(x_d^+)\bar{g}_{t-1}, \\ \bar{y}|_{[t-L_p,t-1]} &= \mathcal{H}|_{[1,L_p]}(y_d)\bar{g}_{t-1}, \end{aligned} \tag{22}$$

then g_t, for $t > 0$, satisfies the following recurrence relation

$$g_t = Eg_{t-1} + Gf|_{L_t} + \beta_t, \;\; f|_{L_t} := \begin{bmatrix} u|_{L_t} \\ x|_{[t-L_p+1,t]} \\ y|_{[t-L_p,t-1]} \end{bmatrix},$$

$$\mathcal{H}_L^l := \begin{bmatrix} \mathcal{H}_L(u_d) \\ \mathcal{H}|_{[l,L_p+l-1]}(x_d^+) \\ \mathcal{H}|_{[l,L_p+l-1]}(y_d) \end{bmatrix}, \;\; l \in \mathbb{N},$$

$$\mathcal{G}_L^1 := (\mathcal{H}_L^1)^\dagger, \text{ with } l = 1, \; E := \mathcal{G}_L^1 \begin{bmatrix} 0 \\ \mathcal{H}|_{[1,L_p]}(x_d^+) \\ \mathcal{H}|_{[1,L_p]}(y_d) \end{bmatrix}, \tag{23}$$

$$G := \mathcal{G}_L^1 \begin{bmatrix} I_{mL} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; \beta_t := \mathcal{G}_L^1 \begin{bmatrix} 0 \\ \mathcal{H}|_{[1,L_p]}(x_d^+)\delta_t \\ \mathcal{H}|_{[1,L_p]}(y_d)\delta_t \end{bmatrix}. \tag{24}$$

Proof: From Theorem 8, we have

$$z|_{L_t} = \mathcal{H}_L(z_d)\bar{g}_t + \kappa|_{L_t}. \tag{25}$$

Since the L-long trajectories are decomposed into past and future values with respect to the current time t, using (20), we can write (25) as

$$\overbrace{\begin{bmatrix} u|_{L_t} \\ x|_{[t-L_p+1,t]} \\ y|_{[t-L_p,t-1]} \end{bmatrix}}^{=:f|_{L_t}} = \overbrace{\begin{bmatrix} \mathcal{H}_L(u_d) \\ \mathcal{H}|_{[0,L_p-1]}(x_d^+) \\ \mathcal{H}|_{[0,L_p-1]}(y_d) \end{bmatrix}}^{=:\mathcal{H}_L^0} [\bar{g}_t + w_t] \tag{26}$$

$$x|_{[t+1,t+L_f+1]} = \mathcal{H}|_{[L_p,L_p+L_f]}(x_d^+)g_t \tag{27}$$

$$y|_{[t,t+L_f]} = \mathcal{H}|_{[L_p,L_p+L_f]}(y_d)g_t, \tag{28}$$

Now taking expectation, we get

$$\bar{f}|_{L_t} = \mathcal{H}_L^0\bar{g}_t. \tag{29}$$

Taking $\mathcal{G}_L^0 := (\mathcal{H}_L^0)^\dagger$ (pseudo-inverse of $\mathcal{H}_L^0$), we can solve (29) to obtain an estimate for $\bar{g}_t$ given as

$$\bar{g}_t = \mathcal{G}_L^0\bar{f}|_{L_t}.$$

Since $f|_{L_t} = \mathcal{H}_L^0\bar{g}_t + \mathcal{H}_L^0 w_t$, we obtain the noisy estimate

$$\begin{aligned} g_t &= \mathcal{G}_L^0\bar{f}|_{L_t} + \mathcal{G}_L^0\mathcal{H}_L^0 w_t \\ &= \mathcal{G}_L^0(\bar{f}|_{L_t} + \mathcal{H}_L^0 w_t). \end{aligned} \tag{30}$$

As per (22), since $\bar{x}|_{[t-L_p+1,t]}$ and $\bar{y}|_{[t-L_p,t-1]}$ are obtained from g_{t-1} and block-rows 1 to L_p of $H_L(z_d)$, we can plug it in (30) to obtain a recurrence relation for g_t. To this end, first we rewrite $\bar{f}|_{L_t}$ (i.e. $\mathbb{E}[f|_{L_t}]$) as

$$\begin{aligned} \bar{f}|_{L_t} &= \begin{bmatrix} I_{mL} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \bar{f}|_{L_t} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & I_{nL_p} & 0 \\ 0 & 0 & I_{pL_p} \end{bmatrix} \bar{f}|_{L_t} \\ \bar{f}|_{L_t} &= \begin{bmatrix} \bar{u}|_{L_t} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{H}|_{[1,L_p]}(x_d^+) \\ \mathcal{H}|_{[1,L_p]}(y_d) \end{bmatrix} \bar{g}_{t-1}. \end{aligned} \tag{31}$$

Plugging (31) into (30), we get

$$g_t = \mathcal{G}_L^1 \begin{bmatrix} \bar{u}|_{L_t} \\ 0 \\ 0 \end{bmatrix} + \mathcal{G}_L^1 \begin{bmatrix} 0 \\ \mathcal{H}|_{[1,L_p]}(x_d^+) \\ \mathcal{H}|_{[1,L_p]}(y_d) \end{bmatrix} \bar{g}_{t-1} + \mathcal{G}_L^1\mathcal{H}_L^1 w_t. \tag{32}$$

Note that the above relation (32) has the terms $\mathcal{G}_L^1$ and $\mathcal{H}_L^1$ instead of $\mathcal{G}_L^0$ and $\mathcal{H}_L^0$ as in (30). This change in superscript from 0 to 1 simply indicates the number of shifts in the rows in the corresponding Hankel matrix, i.e. from $\mathcal{H}|_{[0,L_p-1]}$ to $\mathcal{H}|_{[1,L_p]}$, to obtain the $x|_{[t-L_p+1,t]}$ and $y|_{[t-L_p,t-1]}$ as an estimate by using g_{t-1}.

Now invoking (21) and writing $w_t = w_{t-1} + \delta_t$, we get

$$\begin{aligned} g_t &= \mathcal{G}_L^1 \begin{bmatrix} \bar{u}|_{L_t} \\ 0 \\ 0 \end{bmatrix} + \mathcal{G}_L^1 \begin{bmatrix} 0 \\ \mathcal{H}|_{[1,L_p]}(x_d^+) \\ \mathcal{H}|_{[1,L_p]}(y_d) \end{bmatrix} \bar{g}_{t-1} \\ &\qquad + \mathcal{G}_L^1\mathcal{H}_L^1 w_{t-1} + \mathcal{G}_L^1\mathcal{H}_L^1\delta_t \\ &= \mathcal{G}_L^1 \begin{bmatrix} \bar{u}|_{L_t} \\ 0 \\ 0 \end{bmatrix} + \mathcal{G}_L^1 \begin{bmatrix} 0 \\ \mathcal{H}|_{[1,L_p]}(x_d^+) \\ \mathcal{H}|_{[1,L_p]}(y_d) \end{bmatrix} \bar{g}_{t-1} \\ &\quad + \mathcal{G}_L^1 \begin{bmatrix} 0 \\ \mathcal{H}|_{[1,L_p]}(x_d^+) \\ \mathcal{H}|_{[1,L_p]}(y_d) \end{bmatrix} w_{t-1} + \mathcal{G}_L^1 \begin{bmatrix} \mathcal{H}_L(u_d) \\ 0 \\ 0 \end{bmatrix} w_{t-1} \\ &\qquad + \mathcal{G}_L^1\mathcal{H}_L^1\delta_t \end{aligned} \tag{33}$$

Define

$$\begin{aligned} \delta^u|_{L_t} &:= \mathcal{H}_L(u_d)w_{t-1} + \mathcal{H}_L(u_d)\delta_t \\ &= \mathcal{H}_L(u_d)w_t \\ \delta^x|_{L_p} &:= \mathcal{H}|_{[1,L_p]}(x_d^+)\delta_t \\ \delta^y|_{L_p} &:= \mathcal{H}|_{[1,L_p]}(y_d)\delta_t. \end{aligned} \tag{34}$$

Then (33) is rewritten as

$$g_t = \mathcal{G}_L^1 \begin{bmatrix} 0 \\ \mathcal{H}|_{[1,L_p]}(x_d^+) \\ \mathcal{H}|_{[1,L_p]}(y_d) \end{bmatrix} g_{t-1} + \mathcal{G}_L^1 \begin{bmatrix} \bar{u}|_{L_t} \\ 0 \\ 0 \end{bmatrix} + \mathcal{G}_L^1 \begin{bmatrix} \delta^u|_{L_t} \\ \delta^x|_{L_p} \\ \delta^y|_{L_p} \end{bmatrix}.$$

Since $u|_{L_t} = \bar{u}|_{L_t} + \mathcal{H}_L(u_d)w_t$ (using (20)), we further obtain that

$$g_t = \underbrace{\mathcal{G}_L^1 \begin{bmatrix} 0 \\ \mathcal{H}|_{[1,L_p]}(x_d^+) \\ \mathcal{H}|_{[1,L_p]}(y_d) \end{bmatrix}}_{=:E} g_{t-1} + \underbrace{\mathcal{G}_L^1 \begin{bmatrix} u|_{L_t} \\ 0 \\ 0 \end{bmatrix}}_{=:Gf|_{L_t}} + \underbrace{\mathcal{G}_L^1 \begin{bmatrix} 0 \\ \delta^x|_{L_p} \\ \delta^y|_{L_p} \end{bmatrix}}_{=:\beta_t}. \tag{35}$$

Thus, we can compactly write (35) as

$$g_t = Eg_{t-1} + Gf|_{L_t} + \beta_t. \tag{36}$$

This completes the proof. ■

Combining (36) with the output and state equations (28) and (27) respectively, we get

$$\begin{aligned} g_t &= Eg_{t-1} + Gf|_{L_t} + \beta_t \\ x|_{[t+1,t+L_f+1]} &= \mathcal{H}|_{[L_p,L_p+L_f]}(x_d^+)g_t \\ y|_{[t,t+L_f]} &= \mathcal{H}|_{[L_p,L_p+L_f]}(y_d)g_t. \end{aligned} \tag{37}$$

Since the above system of equations has a linear structure with additive Gaussian noise, we can apply the Kalman filter-type recursive method to estimate y and x. To this end, we shall consider additional noise terms for the state and output predictions. Thus defining

$$\begin{aligned} &\zeta_t \sim \mathcal{N}(0, R_t), \qquad \xi_t \sim \mathcal{N}(0, \Lambda_t), \\ &H := \mathcal{H}|_{[L_p,L_p+L_f]}(y_d),\ F := \mathcal{H}|_{[L_p,L_p+L_f]}(x_d^+), \end{aligned}$$

we can rewrite (37) as

$$g_t = Eg_{t-1} + Gf|_{L_t} + \beta_t \tag{38}$$

$$y|_{[t,t+L_f]} = Hg_t + \zeta_t \tag{39}$$

$$x|_{[t+1,t+L_f+1]} = Fg_t + \xi_t. \tag{40}$$

Based on this, we now formulate the data-driven Kalman filter-type recursion for estimating g_t, y_t, and eventually x_t.

Remark 4: Note that based on (18), (20) and (34), we have that $\delta^u|_{L_t} \sim \mathcal{N}(0, \Sigma_{\eta|_{L_t}})$ with $\Sigma_{\eta|_{L_t}} = \mathcal{H}_L(u_d)\Sigma_{w_t}\mathcal{H}_L(u_d)^\top$.

Algorithm 3: Data-driven Kalman state estimation.

Given: $\mathcal{H}_L(z_d)$, $\bar{u}_t \in \mathbb{R}^m$, $g_0 \in \mathbb{R}^{T-L+1}$, $w_0 \sim \mathcal{N}(0, \Sigma_{w_0})$, $\delta_t \sim \mathcal{N}(0, \Sigma_t)$, $\zeta_t \sim \mathcal{N}(0, R_t)$ and $\xi_t \sim \mathcal{N}(0, \Lambda_t)$, the Kalman estimation of (g, y, x) are as follows:

Generate noise matrices:

$$\begin{aligned} w_t &= w_{t-1} + \delta_t \quad \text{(cf. (21))} \\ \kappa|_{L_t} &= \mathcal{H}_L(z_d)w_t \quad \text{(cf. (20))} \\ \Sigma_{\eta|_{L_t}} &= \mathcal{H}_L(u_d)w_t\mathcal{H}_L(u_d)^\top \\ u|_{L_t} &\sim \mathcal{N}(\bar{u}|_{L_t}, \Sigma_{\eta|_{L_t}}) \quad \text{(cf. Remark 4)} \\ Q_t &= G\mathcal{H}_L^1\Sigma_{\delta_t}(G\mathcal{H}_L^1)^\top \quad \text{((cf. (23))} \\ \beta_t &\sim \mathcal{N}(0, Q_t) \quad \text{(cf. (33))} \\ \zeta_t &\sim \mathcal{N}(0, R_t), \quad \xi_t \sim \mathcal{N}(0, \Lambda_t) \end{aligned}$$

Prediction step:

$$\begin{aligned} g_{t|t-1} &= Eg_{t-1|t-1} + Gf|_{L_t} \\ P_{t|t-1} &= \mathbb{E}[(g_t - g_{t-1|t-1})(g_t - g_{t-1|t-1})^\top] \\ &= EP_{t-1|t-1}E^\top + G\Sigma_{u|_{L_t}}G^\top + Q_t \\ y_{t|t-1} &= Hg_{t|t-1}, \quad x_{t|t-1} = Fg_{t|t-1} \end{aligned}$$

Measurement step:

$$\begin{aligned} y_t &\in \mathcal{B}(\kappa|_{L_t}) \\ r_t &= y_t - y_{t|t-1} \\ S_t &= \mathbb{E}[r_t r_t^\top] = HP_{t|t-1}H^\top + R_t \end{aligned}$$

Update step:

$$\begin{aligned} K_t &= P_{t|t-1}H^\top S_t^{-1} \\ g_{t|t} &= g_{t|t-1} + K_t r_t \\ P_{t|t} &= [I - K_t H]\, P_{t|t-1} \\ y_{t|t} &= Hg_{t|t}, \quad x_{t|t} = Fg_{t|t}. \end{aligned}$$

Note that in the measurement step, only the output y_t is measured, and consequently, the residual (innovation) is obtained solely from y_t. However, if it is also possible to measure the full state x_t, it can also be used, along with y_t, to obtain the innovation vector. However, since the full state can rarely be observed, it represents a special case that we shall not take into consideration any further. Based on this, the terms ξ_t and Λ_t become superfluous and are also ignored in the subsequent discussions.

VII. Numerical Experiments

In this section, we consider the following MIMO system, which illustrates the results presented in this paper.

$$A = \begin{bmatrix} 0.2 & 0.05 & 0 \\ -0.05 & -0.1 & 0.035 \\ -0.05 & 0 & 0.1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 0 & -1.3 \\ 0 & 3.1 \end{bmatrix},$$
$$C = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{41}$$

We first consider the noise-free case and show the effectiveness of Algorithm 2. We consider Gaussian distributed random inputs with zero mean and unit variance of length $T = 60$ to excite the system and collect the data, denoted by $\text{col}(u_d, y_d)$. Moreover, we consider an input/output trajectory of length $L = 10$, for example, that has some missing entries. Our goal here is twofold: first compute the missing entries of the given input/output trajectory or compute the complete input/output trajectory of length $L = 10$, and then reconstruct the state trajectory based on this recovered input/output trajectory. The missing rows in the input trajectory $u|_L$ and that in the output trajectory $y|_L$ are $M_1 = \{7, 8\}$ and $M_2 = \{5, 6, 7, 8, 9, 10, 11, 12\}$, respectively. Using Algorithm 2, we first compute the complete input/output trajectory $\text{col}(\hat{u}|_L, \hat{y}|_L)$ of length $L = 10$ and then the corresponding state trajectory $x|_L$. The input, output, and state trajectories are shown in Figs. 1, 2, 3. As the data are assumed to be exact (noise-free), the computed input/output/state trajectories perfectly match the true trajectories.

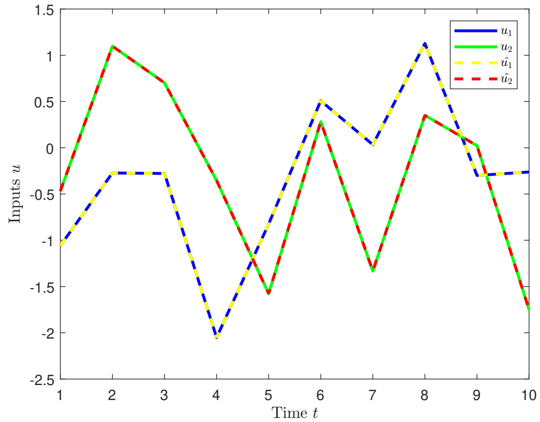

Fig. 1: The recovered input trajectories, using Algorithm 2, coincide with the true input trajectories.

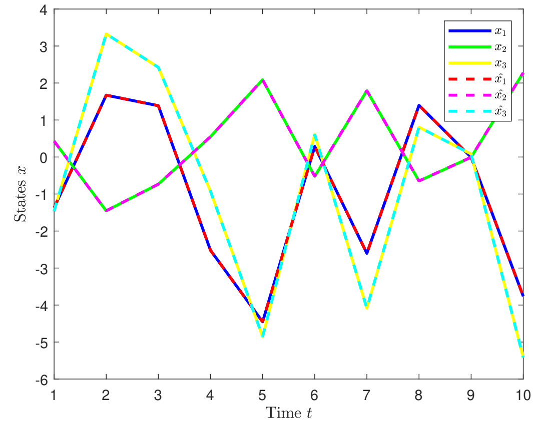

Fig. 3: The reconstructed state trajectories, using Algorithm 2, coincide with the true state trajectories.

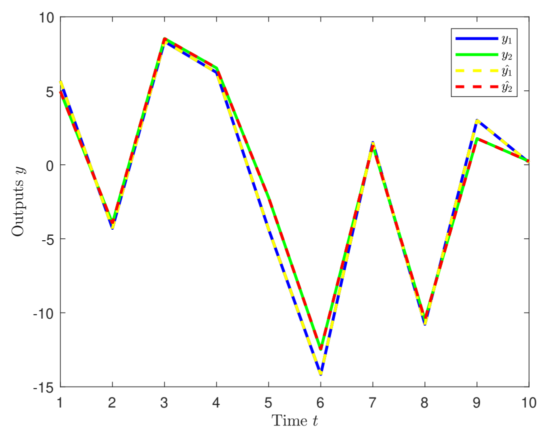

Fig. 2: The recovered output trajectories, using Algorithm 2, coincide with the true output trajectories.

For the case of a noisy system, we shall consider a recursive estimation of the state x_t of the multi-input multi-output system (41). To this end, we consider the use case of reference tracking, wherein $\bar{u}_t$ is generated as per the given reference output $y = [y_1, y_2]^\top$ and the current state x_t. Performing this sequentially in time leads to a data-driven predictive control along with state estimation. This way of testing the algorithm has two-fold advantages: firstly it enables us to validate that the estimated state is sufficiently robust and secondly, it also enables us to validate if the estimated state is robust enough to be used by the data-driven control synthesis program and remains stable for long term reference tracking problem. Thus validating its effectiveness for real-world applications. The initial data and the covariance matrices for input and measurement noise are taken as follows:

$$\Sigma_{w_0} = \Sigma_{\delta_t} = 0.01I, \quad R_t = 0.05I, \quad \Lambda_t = 0, \ \forall t \geq 0$$
$$x_0 = [1, 0, 0]^\top, \quad u_0 = [0, 0], \quad g_0 = \mathcal{G}_L^0 f|_{[0,L_p]}.$$

Moreover, we take $L_p = 3,\ L_f = 6,\ L = 10,\ T = 60$. Since x_t is not used in the residual computation, ξ_t and Λ_t terms are ignored. Based on this, we now use Algorithm 3 to estimate the output $y_{t|t}$ and $x_{t|t}$, in combination with data-driven predictive control for generating reference tracking control $\bar{u}_t$. The reference output y_t is taken as a connected closed curve as shown by the dashed blue curve in the upper left plot of Fig. 4. Since the reference output forms a loop, we run the algorithm for $t = [1, 250]$, which indicates the real evolution of time. The obtained estimates are shown as a dashed-dotted green curve in Fig. 4. At first glance, we see that both the estimates state $x_{t|t}$ and output $y_{t|t}$ are robust and qualitatively follow the true curve profile. The box-plot for pathwise errors $e_{x_t} := \|x_t - x_{t|t}\|_2$ and $e_{y_t} := \|y_t - y_{t|t}\|_2$ of the state and output estimates, respectively, for varying degree of measurement noise, parameterized by γ, are as shown in Fig. 5. The box-plot for each γ was obtained by performing 100 Monte-Carlo simulations. From the plot, we see that both state and output errors have their median in a tolerable range (dashed black line at the center of the box). When the measurement noise covariance R_t is approximately the same as that of δ_t acting on g_t, i.e. $R_t = \gamma \Sigma_{\delta_t}$ with $\gamma \leq 5$, the inter quartile range (IQR) is fairly narrow and the median values of e_{x_t} and e_{y_t} are below 0.05 and 0.075, respectively. As the intensity of measurement noise R_t deviates more and more from that of δ_t, (i.e. for $\gamma \geq 5$), the IQR stretches indicating an increase in the mean error and its variance as indicated in Fig. 5 for $\gamma \in \{10, 25\}$.

VIII. Conclusions

We have stated and proven a generalization of the fundamental lemma due to Willems and co-workers, which includes also the state trajectories. Although the proof is a simple extension of the proof given in [17], the result has been particularly useful in estimating the states of the system. We have used this result to estimate the states in deterministic and stochastic settings. In the deterministic case, we have developed data-driven conditions under which we can reconstruct state trajectories uniquely. Furthermore, we have discussed the case of estimating the states when we have missing input/output data in online and/or offline scenarios.

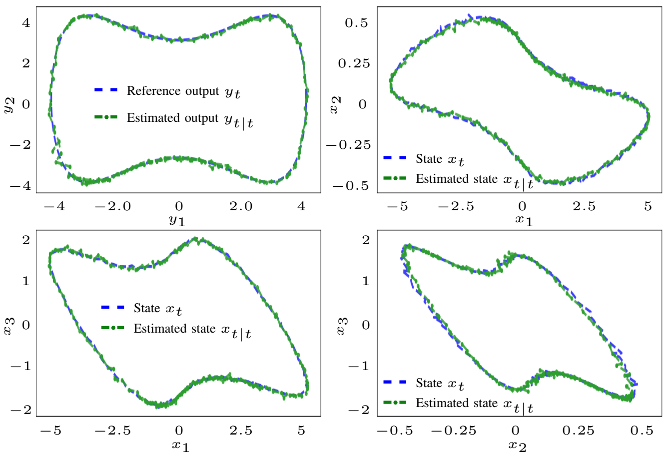

Fig. 4: Output and state estimation by using Algorithm 3. The top left subplot shows the reference output y_t (dashed blue) and estimated output $y_{t|t}$ (dashed-dotted green). The subplots from upper right, bottom left, and bottom right depict the projection of the three-dimensional state onto two-dimensional planes spanned by the coordinate pairs (x_1, x_2), (x_1, x_3), and (x_2, x_3), respectively. Here the true state x_t is denoted in dashed blue while the estimated state $x_{t|t}$ is denoted in dashed-dotted green.

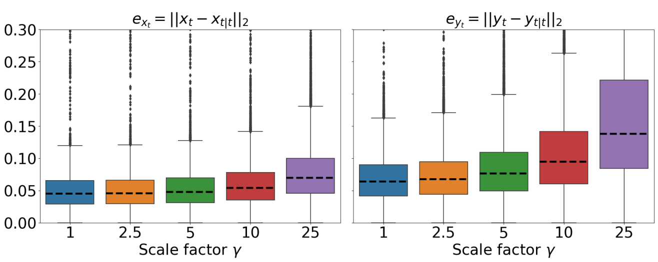

Fig. 5: Box plot for errors, along the trajectory, of the estimated output and state vectors for varying γ, where $R_t = \gamma \Sigma_{\delta_t}$.

In the stochastic case, we have developed a Kalman filter-like algorithm to recursively estimate both states and outputs. Due to space constraints, we have not discussed missing data cases in the stochastic setting. However, we are confident that the results of Section V can be generalized to the case of noisy data. This work has assumed the availability of *offline* state data for estimating the states. It would be interesting to extend the results of this paper to a more natural setting, where only input/output data are available.

References

[1] D. Luenberger, "An introduction to observers," *IEEE Trans. Autom. Control*, vol. 16, no. 6, pp. 596–602, 1971.

[2] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.

[3] J. C. Willems, "The behavioral approach to open and interconnected systems," *IEEE Control Syst. Mag.*, vol. 27, no. 6, pp. 46–99, 2007.

[4] J. C. Willems, P. Rapisarda, I. Markovsky, and B. De Moor, "A note on persistency of excitation," *Syst. Control Lett.*, vol. 54, no. 4, pp. 325–329, 2005.

[5] V. K. Mishra and I. Markovsky, "The set of linear time-invariant unfalsified models with bounded complexity is affine," *IEEE Trans. Autom. Control*, vol. 66, no. 9, pp. 4432–4435, 2020.

[6] I. Markovsky and F. Dörfler, "Identifiability in the behavioral setting," *homepages. vub. ac. be/imarkovs/publications/identifiability. pdf*, 2020.

[7] V. K. Mishra, I. Markovsky, and B. Grossmann, "Data-driven tests for controllability," *IEEE Control Syst. Lett.*, vol. 5, no. 2, pp. 517–522, 2020.

[8] Y. Yu, S. Talebi, H. J. van Waarde, U. Topcu, M. Mesbahi, and B. Açıkmeşe, "On controllability and persistency of excitation in data-driven control: Extensions of Willems' fundamental lemma," in *60th Conf. Decis. Control*. IEEE, 2021, pp. 6485–6490.

[9] H. J. van Waarde, C. D. Persis, M. Camlibel, and P. Tesi, "Willems' fundamental lemma for state-space systems and its extension to multiple datasets," *IEEE Control Syst. Lett.*, vol. 4, pp. 602–607, 2020.

[10] J. G. Rueda-Escobedo and J. Schiffer, "Data-driven internal model control of second-order discrete volterra systems," in *IEEE Conf. Decision Control*. IEEE, 2020, pp. 4572–4579.

[11] V. K. Mishra, I. Markovsky, A. Fazzi, and P. Dreesen, "Data-driven simulation for NARX systems," in *29th Eur. Signal Process. Conf.*, Dublin, Ireland, August 2021, pp. 1–5.

[12] J. Berberich and F. Allgöwer, "A trajectory-based framework for data-driven system analysis and control," in *Eur. Control Conf.* IEEE, 2020, pp. 1365–1370.

[13] I. Markovsky and F. Dörfler, "Behavioral systems theory in data-driven analysis, signal processing, and control," *Annu. Rev. Control*, vol. 52, pp. 42–64, 2021.

[14] I. Markovsky and P. Rapisarda, "Data-driven simulation and control," *Int. J. Control*, vol. 81, no. 12, pp. 1946–1959, 2008.

[15] T. Maupong and P. Rapisarda, "Data-driven control: A behavioral approach," *Syst. Control Lett.*, vol. 101, pp. 37–43, 2017.

[16] A. A. Al Makdah, V. Krishnan, V. Katewa, and F. Pasqualetti, "Behavioral feedback for optimal lqg control," in *61st Conf. Decis. Control*. IEEE, 2022, pp. 4660–4666.

[17] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 909–924, 2019.

[18] J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control: In the shallows of the DeePC," in *Eur. Control Conf.* IEEE, 2019, pp. 307–312.

[19] J. Berberich, J. Köhler, M. A. Muller, and F. Allgower, "Data-driven model predictive control with stability and robustness guarantees," *IEEE Trans. Autom. Control*, 2020.

[20] S. Baros, C.-Y. Chang, G. E. Colon-Reyes, and A. Bernstein, "Online data-enabled predictive control," *arXiv preprint arXiv:2003.03866*, 2020.

[21] S. A. Hiremath, V. K. Mishra, and N. Bajcinca, "Learning based stochastic data-driven predictive control," in *61st Conf. Decis. Control*. IEEE, 2022, pp. 1684–1961.

[22] G. Pan, R. Ou, and T. Faulwasser, "On a stochastic fundamental lemma and its use for data-driven optimal control," *IEEE Trans. Autom. Control*, 2022.

[23] J. Shi, Y. Lian, and C. N. Jones, "Data-driven input reconstruction and experimental validation," *arXiv preprint arXiv:2203.02827*, 2022.

[24] V. K. Mishra, A. Iannelli, and N. Bajcinca, "A data-driven approach to system invertibility and input reconstruction," in *62nd Conf. Decis. Control*. IEEE, 2023, pp. 671–676.

[25] T. M. Wolff, V. G. Lopez, and M. A. Müller, "Robust data-driven moving horizon estimation for linear discrete-time systems," *IEEE Trans. Autom. Control*, 2024.

[26] I. Markovsky, L. Huang, and F. Dörfler, "Data-driven control based on the behavioral approach: From theory to applications in power systems," *IEEE Control Syst.*, 2022.

[27] Y. Yan, J. Bao, and B. Huang, "On approximation of system behavior from large noisy data using statistical properties of measurement noise," *IEEE Trans. Autom. Control*, 2023.

[28] D. Alpago, F. Dörfler, and J. Lygeros, "An extended Kalman filter for data-enabled predictive control," *IEEE Control Systems Letters*, vol. 4, no. 4, pp. 994–999, 2020.

[29] M. S. Turan and G. Ferrari-Trecate, "Data-driven unknown-input observers and state estimation," *IEEE Control Syst. Lett.*, vol. 6, pp. 1424–1429, 2021.

[30] V. K. Mishra, H. J. van Waarde, and N. Bajcinca, "Data-driven criteria for detectability and observer design for LTI systems," in *61st Conf. Decis. Control*. IEEE, 2022, pp. 4846–4852.

[31] H. Van Waarde, J. Eising, H. Trentelman, and K. Camlibel, "Data informativity: a new perspective on data-driven analysis and control," *IEEE Trans. Autom. Control*, 2020.

[32] I. Markovsky and F. Dörfler, "Data-driven dynamic interpolation and approximation," *Automatica*, vol. 135, p. 110008, 2022.