On Controllability and Persistency of Excitation in Data-Driven Control: Extensions of Willems’ Fundamental Lemma

Yue Yu, Shahriar Talebi, Henk J. van Waarde, Ufuk Topcu, Mehran Mesbahi, and Behçet Açıkmeşe

Abstract—Willems’ fundamental lemma asserts that all trajectories of a linear time-invariant (LTI) system can be obtained from a finite number of measured ones, assuming controllability and a persistency of excitation condition hold. We show that these two conditions can be relaxed. First, we prove that the controllability condition can be replaced by a condition on the controllable subspace, unobservable subspace, and a certain subspace associated with the measured trajectories. Second, we prove that the persistency of excitation requirement can be reduced if the degree of certain minimal polynomial is known or tightly bounded. Our results shows that data-driven predictive control using online data is equivalent to model predictive control, even for uncontrollable systems. Moreover, our results significantly reduce the amount of data needed in identifying homogeneous multi-agent systems.

I. INTRODUCTION

Willems’ fundamental lemma provides a data-based parameterization of trajectories generated by linear time invariant (LTI) systems [1]. In particular, consider the LTI system

\[ x_{t+1} = Ax_t + Bu_t, \]
\[ y_t = Cx_t + Du_t, \]

where \( u_t \in \mathbb{R}^m, x_t \in \mathbb{R}^n, y_t \in \mathbb{R}^p \) denote the input, state and output of the system at discrete time \( t \), respectively. If system (1) is controllable, the lemma asserts that every length-\( L \) input-output trajectory of system (1) is a linear combination of a finite number of measured ones. These measured trajectories can be extracted from one single trajectory with persistent excitation of order \( n + L \) [1], or multiple trajectories with collective persistent excitation of order \( n + L \) [2]. By parameterizing trajectories of the system (1) using measured data, the lemma has profound implications in system identification [3], [4], [5], and inspired a series of recent results including data-driven simulation [3], [6], output matching [7], control by interconnection [8], set-invariance control [9], linear quadratic regulation [10], and predictive control [11], [12], [13], [14], [15], [16], [17].

On the other hand, the conditions of controllability and persistency of excitation in Willems’ fundamental lemma are necessary but has not been investigated to depth. The current work aims to address this issue by answering the following two questions. First, with persistency of excitation, to what extent can the linear combinations of finite number of measured input-output trajectories parameterize all possible ones? Second, can the order of persistent excitation be reduced?

Recent results on system identification has shown that, assuming sufficient persistency of excitation, the linear combinations of a finite number of measured input-output trajectories contain any trajectory whose initial state is in the controllable subspace [18], [19]. As we show subsequently, such results only partially answer the first question above: trajectories with initial state outside the controllable subspace can also be contained in the said linear combinations.

We answer the aforementioned questions by introducing extensions of Willems’ fundamental lemma. First, we show that, with sufficient persistency of excitation, any length-\( L \) input-output trajectory whose initial state is in some subspace is a linear combination of a finite number of measured ones. The said subspace is the sum of the controllable subspace, unobservable subspace, and a certain subspace associated with the measured trajectories. Second, we show that the order of persistent excitation required by Willems’ fundamental lemma can be reduced from \( n + L \) to \( \delta_{\min} + L \), where \( \delta_{\min} \) is the degree of the minimal polynomial of the system matrix \( A \). Our first result completes those presented in [18], [19] by showing exactly which trajectories are parameterizable by a finite number of measured ones for an arbitrary LTI systems. Furthermore, this result show that data-driven predictive control using online data is equivalent to model predictive control, not only for controllable systems, as shown in [12], [13], but also for uncontrollable systems. Our second result, compared with those in [2], reduces the amount of data samples used in identifying homogeneous multi-agent systems by one order of magnitude.

The rest of the paper is organized as follows. We first prove an extended Willems’ fundamental lemma and discuss its implications in Section II. We provide ramifications of this extension for a representative set of applications in Section III before providing concluding remarks in Section IV.

Notation: We let \( \mathbb{R} \), and \( \mathbb{N} \) and \( \mathbb{N}_+ \) denote the set of real numbers, non-negative integers, and positive integers, respectively. The image and right kernal of matrix \( M \) is denoted by \( \text{im} M \) and \( \ker M \), respectively. The Moore-Penrose inverse of \( M \) is denoted by \( M^+ \). When applied to subspaces, we let + and \( \times \) denote the sum [20, p.2] and Cartesian product operation [20, p.370], respectively. We let \( \otimes \) denote the Kronecker product. Given a signal
has full row rank.

II. EXTENSIONS OF WILLEMS' FUNDAMENTAL LEMMA

In this section, we introduce extensions of Willems' fundamental lemma. Throughout we let

\[ (u^i_{[0,T^i-1]}, x^i_{[0,T^i-1]}, y^i_{[0,T^i-1]}) \]

denote a length-\( T^i \) (\( T^i \in \mathbb{N}_+ \)) input-state-output trajectory generated by system (1) for all \( i = 1, 2, \ldots, r \), where \( r \in \mathbb{N}_+ \) is the total number of trajectories. We let

\[ \{u^i_{[0,T^i-1]}\}_{i=1}^r, \{x^i_{[0,T^i-1]}\}_{i=1}^r, \{y^i_{[0,T^i-1]}\}_{i=1}^r, \]

denote the set of input, state, and output trajectories, respectively. We will use the following subspaces

\[ \mathcal{R} = \text{im} \left[ B \quad AB \quad \cdots \quad A^{n-1}B \right], \]

\[ \mathcal{O} = \ker \left[ C^\top \left( CA \right)^\top \cdots \left( CA^{n-1} \right)^\top \right], \]

\[ \mathcal{K}[x_0, x_2, \ldots, x_0^r] = \text{im} \left[ X_0 \quad AX_0 \quad \cdots \quad A^{n-1}X_0 \right], \]

where \( X_0 = [x_0 \quad x_2 \quad \cdots \quad x_0^r] \). In particular, \( \mathcal{R} \) is known as the controllable subspace, \( \mathcal{O} \) is known as the unobservable subspace, \( \mathcal{K}[x_0, x_2, \ldots, x_0^r] \) is the controllable subspace with \( B \) replaced by \( X_0 \). We say system (1) is controllable if \( \mathcal{R} = \mathbb{R}^n \). One can verify that \( \mathcal{R}, \mathcal{O} \) and \( \mathcal{K}[x_0, x_2, \ldots, x_0^r] \) are all invariant subspace of matrix \( A \).

We will also use the following definitions that streamline the subsequent analysis.

**Definition 1.** We say a length-L input-output trajectory \( (\bar{u}^i_{[0,L-1]}, \bar{y}^i_{[0,L-1]}) \) with \( L \in \mathbb{N}_+ \) is parameterizable by \( \{u^i_{[0,T^i-1]}, y^i_{[0,T^i-1]}\}_{i=1}^r \) if there exists \( g \in \mathbb{R}^{r \times (n^r - 1)} \) such that

\[ \begin{bmatrix} \bar{u}^i_{[0,L-1]} \\ \bar{y}^i_{[0,L-1]} \end{bmatrix} = \begin{bmatrix} H_L(u^1_{[0,T^i-1]}) & \cdots & H_L(u^r_{[0,T^i-1]}) \\ H_L(y^1_{[0,T^i-1]}) & \cdots & H_L(y^r_{[0,T^i-1]}) \end{bmatrix} g. \]

(4)

As an example, if \( r = 2, L = 2, T^1 = 3, T^2 = 4 \), then equation (4) becomes the following

\[ \begin{bmatrix} \bar{u}_0 \\ \bar{u}_1 \\ \bar{y}_0 \\ \bar{y}_1 \end{bmatrix} = \begin{bmatrix} u_0^1 & u_2^1 & u_0^2 & u_2^2 \\ u_0^2 & u_2^2 & u_0^1 & u_2^1 \\ y_0^1 & y_1^1 & y_0^2 & y_1^2 \\ y_1^1 & y_2^1 & y_1^2 & y_2^2 \end{bmatrix} g. \]

**Definition 2 (Collective persistent excitation [2]).** We say \( \{u^i_{[0,T^i-1]}\}_{i=1}^r \) is collectively persistently exciting of order \( d \in \mathbb{N}_+ \) if \( d \leq T^i \) for all \( i = 1, 2, \ldots, r \) and the mosaic-Hankel matrix, defined as

\[ H_d(u^1_{[0,T^i-1]}), \cdots, H_d(u^r_{[0,T^i-1]}) \],

has full row rank.

If \( \tau = 1 \), then Definition 2 reduces to the traditional notion of persistency of excitation.

The Willems’ fundamental lemma asserts that: if system (1) is controllable and \( \{u^i_{[0,T^i-1]}\}_{i=1}^r \) is collectively persistently exciting of order \( n + L \), then \( \{\bar{u}^i_{[0,L-1]}, \bar{y}^i_{[0,L-1]}\} \) is parameterizable by \( \{u^i_{[0,T^i-1]}, y^i_{[0,T^i-1]}\}_{i=1}^r \) if and only if it is an input-output trajectory of system (1) [1], [2]. As our main contribution, the following theorem shows that not only this lemma can be extended to an arbitrary LTI system, but also the required order of collective persistent excitation can be reduced from \( n + L \) to \( \delta_{\text{min}} + L \), where \( \delta_{\text{min}} \) is the degree of the minimal polynomial of matrix \( A \).

**Theorem 1.** Let \( \delta_{\text{min}} \) be the degree of the minimal polynomial of matrix \( A \), and \( \delta \in \mathbb{N}_+ \) with \( \delta \geq \delta_{\text{min}} \). Let \( \{u^i_{[0,T^i-1]}, x^i_{[0,T^i-1]}, y^i_{[0,T^i-1]}\}_{i=1}^r \) be the set of input-state-output trajectories generated by system (1). If \( \{u^i_{[0,T^i-1]}\}_{i=1}^r \) is collectively persistently exciting of order \( \delta + L \), then

\[
\begin{bmatrix}
H_1(x^1_{[0,T^i-1]}) & \cdots & H_1(x^r_{[0,T^i-1]}) \\
H_L(x^1_{[0,T^i-1]}) & \cdots & H_L(x^r_{[0,T^i-1]}) \\
& \cdots & \\
H_L(x^1_{[0,T^i-1]}) & \cdots & H_L(x^r_{[0,T^i-1]})
\end{bmatrix}
\]

(6)

\[
= (\mathcal{R} + \mathcal{K}[x_0, x_2, \ldots, x_0^r]) \times \mathbb{R}^{nL}.
\]

Further, \( \{\bar{u}^i_{[0,L-1]}, \bar{y}^i_{[0,L-1]}\} \) is parameterizable by \( \{u^i_{[0,T^i-1]}, y^i_{[0,T^i-1]}\}_{i=1}^r \) if and only if there exists a state trajectory \( \bar{x}_{[0,L-1]} \) with

\[ \bar{x}_0 \in \mathcal{R} + \mathcal{O} + \mathcal{K}[x_0, x_2, \ldots, x_0^r], \]

such that \( \{\bar{u}^i_{[0,L-1]}, \bar{x}^i_{[0,L-1]}, \bar{y}^i_{[0,L-1]}\} \) is an input-state-output trajectory of system (1).

**Proof.** See the Appendix.

**Remark 1.** The equality in (6) generalizes [18, Lem. 2] by proving stronger results using weaker assumptions. Particularly, the assumption of \( n + L \) order of persistently excitation in [18, Lem. 2] is reduced to \( \delta + L \) with \( \delta \geq \delta_{\text{min}} \), and the controllable subspace \( \mathcal{R} \) used in [18, Lem. 2] is extended to its superset \( \mathcal{R} + \mathcal{O} + \mathcal{K}[x_0, x_2, \ldots, x_0^r] \).

**Remark 2.** Theorem 1 generalizes [2, Thm. 2] by proving the same results using weaker assumptions. Particularly, to ensure all input-output trajectories generated by system (1) are parameterizable by a finite number of them, [2, Thm. 2] assumes \( \mathcal{R} = \mathbb{R}^n \). In comparison, using Theorem 1 one only need to assume that \( \mathcal{R} + \mathcal{O} + \mathcal{K}[x_0, x_2, \ldots, x_0^r] = \mathbb{R}^n \) to ensure the same results.

The first statement in Theorem 1 leads to a new result in system identification, summarized by the following corollary.

**Corollary 1.** Let \( \{u^i_{[0,T^i-1]}, x^i_{[0,T^i-1]}\}_{i=1}^r \) be input-state-trajectories generated by system (1), and input sequences \( \{u^i_{[0,T^i-1]}\}_{i=1}^r \) are collectively persistently exciting of order \( \delta + 1 \) with \( \delta \geq \delta_{\text{min}} \), where \( \delta_{\text{min}} \) is the degree of the minimal polynomial of \( A \). Define \( [A \quad B] := X \Sigma^T \), where

\[
X = \begin{bmatrix}
H_1(x^1_{[1,T^1]}) & \cdots & H_1(x^r_{[1,T^r]})
\end{bmatrix},
\]

(8a)

The minimal polynomial of a matrix \( A \) is the unique monic polynomial of minimum degree that annihilates matrix \( A \) [20, Def. 3.3.2].
Fact the minimal polynomial of $A$ only depends on the degree of the minimal polynomial of $A$. Corollary 1 reduces to Theorem 1 in [10].

If $\delta = n$ and system $[A]$ is controllable, then Corollary [7] reduces to Theorem 1 in [10].

Another implication of Theorem 1 is that the order of persistent excitation required by trajectory parameterization only depends on the degree of the minimal polynomial of matrix $A$ in $[A]$, instead of its dimension. In general, it is difficult to establish a bound of the degree of the minimal polynomial of a matrix tighter than its dimension. However, the following corollary shows an exception example; its usefulness will be illustrated later in Section III-B.

**Corollary 2.** If there exists $\tilde{A} \in \mathbb{R}^{n \times n}$ such that $A = I_N \otimes \tilde{A}$, then Theorem 1 holds with $\delta = \frac{n}{N}$.

**Proof.** The results directly follow from Theorem 1 and the fact the minimal polynomial of $A$ is the same as the one of $\tilde{A}$, which has degree at most $\frac{n}{N}$. \hfill $\square$

### III. APPLICATIONS

In this section, we provide two examples that illustrate distinct implications of Theorem 1.

#### A. Online data-driven predictive control

Model predictive control (MPC) provides an effective strategy for systems with physical and operational constraints [21], [22]. In particular, consider system $[1a]$. At each sampling time $t$, MPC solves the following optimization to obtain the input $u_t$

$$
\text{minimize}_{t_i, t_i+1, \ldots, t_i+t-1} \sum_{k=t}^{t+L-1} \left( \| x_k - r_k \|_Q^2 + \| u_k \|_R^2 \right) \\
\text{subject to} \quad \begin{align*}
\dot{x}_{k+1} &= A\hat{x}_k + Bu_k, \quad \hat{x}_t = 0, \\
\hat{x}_k &\in X, \quad \hat{\theta}_k \in U, \quad k = t, \ldots, t + L - 1,
\end{align*}
$$

(9)

where $\hat{x}_k \in \mathbb{R}^n$ is the current state and $L \in \mathbb{N}$ is the planning horizon. Closed convex sets $X \subset \mathbb{R}^n$ and $U \subset \mathbb{R}^m$ describe feasible states and inputs, respectively. Symmetric positive semi-definite weighting matrices $Q \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{m \times m}$, together with reference state trajectory $r^\ast(t, t+L-1)$, define the quadratic tracking cost function.

Recently, [12], [13] proposed data-driven predictive control (DPC) that replaces optimization (9) with

$$
\text{minimize}_{\tilde{t}_i, \tilde{t}_i+1, \ldots, \tilde{t}_i+L-1} \sum_{k=\tilde{t}}^{\tilde{t}+L-1} \left( \| x_k - r_k \|_Q^2 + \| u_k \|_R^2 \right) \\
\text{subject to} \quad \begin{align*}
\hat{x}_{k+1} &= A\hat{x}_k + Bu_k, \quad \hat{x}_\tilde{t} = \hat{x}_t, \\
\hat{x}_k &\in X, \quad \hat{\theta}_k \in U, \quad k = \tilde{t}, \ldots, \tilde{t} + L - 1,
\end{align*}
$$

(10)

where $\tilde{t}_i \in \mathbb{R}$ is an input-state trajectory of system $[1a]$ and generated offline. If $u^\ast(t_0, T-1)$ is persistently exciting of order $n + L$ and system $[1a]$ is controllable, Willems’ fundamental lemma guarantees that optimization (10) is equivalent to the one in (9).

However, testing the controllability of system (1a) using its input-state data $(u^0(t_0, T-1), x^0(t_0, T-1))$ is expensive: its computation time scales cubically with $T$ [23], [18]. Further, if system $[1a]$ is uncontrollable, then Willems’ fundamental lemma provides no guarantee on the equivalence between optimization (10) and (9).

On the other hand, Theorem 1 shows that the assumption of $(A, B)$ being controllable can be replaced by $\hat{x}_t \in \mathcal{R} + K[x_0]$ in $[1a]$. In particular, if there exists an input sequence $\hat{u}_0, \hat{u}_1, \ldots, \hat{u}_{k-1} \in \mathbb{R}^m$ such that

$$
\hat{x}_t = A^k x_0 + \sum_{k=0}^{k-1} A^{k-j-1} B\hat{u}_j.
$$

(11)

Using the Cayley-Hamilton theorem, one can verify that $x_{t+1}$ remains persistently exciting of order $n + L$.

#### B. Identification of homogeneous multi-agent systems

Consider a network of $N$ agents with the same LTI dynamics [24]. Further, agent $i$ can measure the state of agent $j$ in a local coordinate system if $(i, j)$ is an edge of a directed
a sampling time of 0.1s such that the system dynamics is given by (14) with

\[ A = I_N \otimes \bar{A}, B = I_N \otimes \bar{B}, C = E \otimes I_n, D = 0_{\bar{m} \times Nn}, \]

where \( \bar{A} \in \mathbb{R}^{\bar{n} \times \bar{n}} \) and \( \bar{B} \in \mathbb{R}^{\bar{n} \times \bar{m}} \) describe the dynamics of an individual agent. Each row of matrix \( E \in \mathbb{R}^{M \times N} \) is indexed by an directed edge, i.e., an edge with an head and a tail, in graph \( G \): the \( i \)-th entry in each row is “1” if node \( i \) is the head of the corresponding edge, “-1” if it is the tail, and “0” otherwise. We assume that \( \bar{B} \) is a non-zero matrix and \( (\bar{A}, \bar{B}) \) is controllable.

If at least one non-zero entry in matrix \( E \) is known, then system matrices in (14) can be computed using the following Markov parameters [25, Sec. 3.4.4]

\[
M_k = CA^{k-1}B + D = (E \otimes I_n)(I_N \otimes \bar{A})^{-1} = E \otimes (\bar{A}^{-1}B), \quad \forall k = 1, 2, \ldots, \bar{n} + 1. \tag{15}
\]

In particular, let \((M_k)_{ij}\) denote the \(ij\)-th \(\bar{n} \times \bar{m}\) block of \(M_k\).

If we know \(E_{ij} = 1\) (the case of “-1” is similar), then (15) implies \((M_k)_{ij} = \bar{A}^{k-1}B\). For example, if \(E = \begin{bmatrix} 1 & -1 \end{bmatrix}\), then (15) says \(M_k = \begin{bmatrix} \bar{A}^{k-1}B & -\bar{A}^{k-1}B \end{bmatrix}\). Hence given the Markov parameters (15) and that \(E_{ij} = 1\), we know \(E_{kl}\) is “1” if \((M_k)_{ij} = (M_k)_{ij}\), “-1” if \((M_k)_{ij} = -(M_k)_{kl}\), and “0” otherwise. Further, \(\bar{B} = (M_k)_{ij}\) and \(\bar{A}\) is the unique solution to the following linear equation:

\[
\bar{A} \begin{bmatrix} (M_1)_{ij} & \cdots & (M_{\bar{n}})_{ij} \end{bmatrix} = \begin{bmatrix} (M_{\bar{n}+1})_{ij} \end{bmatrix}. \tag{16}
\]

Therefore, given at least one non-zero entry in matrix \(E\), in order to compute the system matrices in (14), it suffices to know the Markov parameters (15). These parameters can be computed using Corollary 2 via a data-driven simulation procedure [3], [7]. We include the detailed steps of this procedure in the Appendix for completeness.

In numerical simulations, we consider the homogeneous multi-agent system used in [24, Example 3], discretized with a sampling time of 0.1s such that the system dynamics is given by (14) where

\[
\bar{A} = \begin{bmatrix} 0.9964 & 0.0026 & -0.0004 & -0.0460 & 0.0045 & 0.9037 & -0.0128 & -0.3834 & 0.0098 & 0.0339 & 0.0838 & 0.1302 & 0.0005 & 0.0017 & 0.0968 & 1.0067 \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} 0.0445 & 0.0167 & 0.3407 & -0.7249 & -0.5278 & 0.4214 & 0.2068 & 0.0215 \end{bmatrix}. \tag{5}
\]

In addition, we let \( E = \begin{bmatrix} 1_{N-1} & -I_{N-1} \end{bmatrix} \), where \(1_{N-1} \in \mathbb{R}^{N-1}\) is the vector of all 1’s.

We compare Corollary 2 against the results in [2, Thm. 2] in terms of the least amount of input-output data needed to compute matrices in (14). In particular, since matrix \(\bar{A}\) has spectrum radius 1.03, we use use input-output trajectories \(\{u_{t}^{i}[0,T-1], y_{t}^{i}[0,T-1]\}_{t=1}^{N}\) with relatively short length \(T = 120\) to avoid numerical instability [2]. The entries in \(\{u_{t}^{i}[0,T-1]\}_{t=1}^{N}\) are sampled uniformly from \([-0.1, 0.1]\). Using Corollary 2, the data-driven simulation procedure requires \(\{u_{t}^{i}[0,T-1]\}_{t=1}^{N}\) to be collectively persistently exciting of order \((N + 1)\bar{n} + 1\). In other words, matrix (5) with \(d = (N + 1)\bar{n} + 1\) has full row rank, hence it must have at least as many columns as rows, i.e.,

\[
\tau \geq \frac{(N+1)\bar{n}+1N\bar{m}}{T - (N+1)\bar{n}} = \frac{8N^2 + 10N}{16}.
\]

In comparison, if we use [2, Thm. 2] instead of Corollary 2, we need \(\{u_{t}^{i}[0,T-1]\}_{t=1}^{N}\) to be collectively persistently exciting of order \(2N\bar{n} + 1\) (see [2, Sec. IV-A]). In other words, matrix (5) with \(d = 2N\bar{n} + 1\) has full row rank, which implies

\[
\tau \geq \frac{(2N+1)\bar{n}+1N\bar{m}}{T - 2N\bar{n}} = \frac{16N^2 + 2N}{120}.
\]

In Fig. 2 we show the minimum number of input-output trajectories required to compute matrices in (14) in numerical simulations. The results tightly match the aforementioned two lower bounds, and the number of trajectories required by Corollary 2 is one order of magnitude less than that of [2, Thm. 1] when \(N = 14\).

\[ \text{Fig. 2. Minimum number of length } T = 120 \text{ input-output trajectories required to identify system } (14). \]

\[ \text{IV. CONCLUSIONS} \]

We introduced an extension of Willems’ fundamental lemma by relaxing the controllability and persistency of excitation assumptions. We demonstrate the usefulness of our results in the context of DPC and identification of homogeneous multi-agent system. Future directions include generalizations to noisy data and nonlinear systems.

\[ \text{APPENDIX} \]

\[ \text{Proof of Theorem 2} \]

We start by proving the first statement using a double inclusion argument. It is trivial to show that the left hand side of (5) is included in its right...
hand side. To show the other direction, we show that the left kernel of matrix
\[
\begin{bmatrix}
H_1(x_{0,T'-L}) & \cdots & H_1(x_{0,T'-L}) \\
H_L(u_{0,T'-1}) & \cdots & H_L(u_{0,T'-L})
\end{bmatrix}
\]
is orthogonal to \((R + K[x_0^T, x_0^T, \ldots, x_0^T]) \times \mathbb{R}^{mL}\). To this end, let
\[
v^\top = \begin{bmatrix}
\xi^\top \\
\eta_1^\top \\
\eta_2^\top \\
\vdots \\
\eta^\top_L
\end{bmatrix}
\]
be an arbitrary row vector in the left kernel of matrix \((17)\), where \(\xi \in \mathbb{R}^n, \eta_1, \eta_2, \ldots, \eta_L \in \mathbb{R}^m\). Since \(\delta \geq \delta_{\text{min}}\), using [20, Def. 3.3.2] we know there exists \(\alpha_{0k}, \alpha_{1,k}, \ldots, \alpha_{\delta-1,k} \in \mathbb{R}\) such that
\[
A^k + \sum_{j=0}^{\delta-1} \alpha_{jk} A^j = 0_{n \times n}, \quad \forall k = \delta, \delta + 1, \ldots,
\]
The above equation implies that \(A^k B = -\sum_{j=0}^{\delta-1} \alpha_{jk} A^j B\) and \(A^{k-j} = -\sum_{j=0}^{\delta-1} \alpha_{jk} A^j x_0^T\) for all \(k = \delta, \delta + 1, \ldots, n - 1\) and \(i = 1, \ldots, \tau\). Therefore in order to show the left kernel of matrix \((17)\) is orthogonal to \((R + K[x_0^T, x_0^T, \ldots, x_0^T]) \times \mathbb{R}^{mL}\), it suffices to show the following
\[
\begin{align*}
\eta_1^\top &= \cdots = \eta^\top_L = 0_m, \\
\xi^\top \quad B &= 0_m, \\
\xi^\top \quad A^i B &= 0_m, \\
\xi^\top \quad x_0^T &= 0_m,
\end{align*}
\]
for \(j = 1, \ldots, \delta\). Since \(v^\top \) is in the left kernel of matrix \((17)\), using \((1)\) one can verify that \(w_0^\top, w_1^\top, \ldots, w_\delta^\top\) are in the left kernel of
\[
\begin{bmatrix}
H_1(x_{0,T'-\delta-L}) & \cdots & H_1(x_{0,T'-\delta-L}) \\
H_{\delta+L}(u_{0,T'-1}) & \cdots & H_{\delta+L}(u_{0,T'-1})
\end{bmatrix}.
\]
Let \(\alpha_{\delta k} = 1\) and \(k = \delta\) in \((19)\), we have \(0_{n \times n} = \sum_{j=0}^{\delta} \alpha_{jk} A^j I\). Hence
\[
\sum_{j=0}^{\delta} \alpha_{jk} w_j^\top = \sum_{j=0}^{\delta} \alpha_{jk} \xi^\top A^j r^\top = [0_n \quad r^\top],
\]
for some vector \(r \in \mathbb{R}^{m(\delta + L)}\). Since row vectors \(w_0^\top, w_1^\top, \ldots, w_\delta^\top\) are in the left kernel of matrix \((17)\), equation \((22)\) implies that \(r^\top \) is in the left kernel of matrix
\[
\begin{bmatrix}
H_{\delta+L}(u_{0,T'-1}) & \cdots & H_{\delta+L}(u_{0,T'-1})
\end{bmatrix}.
\]
Since \(\{u_{0,T'-1}\}_{i=1}^T\) is collectively persistently exciting of order \(\delta + L\), matrix \((23)\) has full row rank. Therefore
\[
r = 0_{m(\delta + L)}.
\]
Next, since \((20a)\) holds, the first \(m \delta\) entries in \(r\) are
\[
\begin{bmatrix}
\sum_{j=1}^{\delta} A^j \xi^\top A^j & \sum_{j=2}^{\delta} A^j \xi^\top A^{j-1} \xi^\top A^j & \cdots & A^2 \xi^\top A \xi^\top A^j & \xi^\top A^j
\end{bmatrix}^\top.
\]
By combining this with \((24)\) we get that
\[
0_m = \sum_{j=k}^{\delta} \alpha_{\delta k} \xi^\top A^j k B, \quad \forall k = 1, \ldots, \delta.
\]
Since \(\alpha_{\delta k} = 1\), considering \(k = \delta\) in \((25)\) implies that \(\xi^\top A\xi^\top = 0_m\). By repeating a similar induction we can prove that \((20a)\) holds.

Further, by using \((19)\) and \((20b)\) we can show that \(\xi^\top A\xi^\top = 0_m\) for all \(k \geq 1\). Combining this together with the fact that row vector \((18)\) is in the left kernel of matrix \((17)\) and that \((20a)\) also holds, it follows that
\[
0 = \xi^\top x_0^T = \xi^\top (A^k \xi_0 + \sum_{j=0}^{k-1} A^{k-j-1} B x_0),
\]
where \(x_0^T = 0_m \quad T - 1, \quad i = 1, \ldots, \tau\).

Since \(T \geq \delta + L\) for \(i = 1, \ldots, \tau\) by assumption, we conclude that \((20a)\) holds.

We now prove the second statement. Given the input, state, and output trajectories in \((3)\), suppose \((0)\) holds. Let
\[
x_0 = \begin{bmatrix} H_1(x_{0,T'-L}) & \cdots & H_1(x_{0,T'-L}) \end{bmatrix} g,
\]
\[
x_{i+1} = A x_i + B u_i, \quad 0 \leq i \leq L - 2.
\]
Then from \((6)\) we know \(x_0\) satisfy \((7)\), and one can verify that \((u_{[0,L-1]}, \tilde{x}_{[0,L-1]}, y_{[0,L-1]}\) is indeed an input-state-output trajectory of system \((1)\).

Conversely, let \((u_{[0,L-1]}, \tilde{x}_{[0,L-1]}, y_{[0,L-1]}\) be an input-state-output trajectory of system \((1)\) with \(x_0 \in R + K[x_0^T, x_0^T, \ldots, x_0^T]\). Then there exists
\[
\tilde{x}_0^\delta \in R + K[x_0^T, x_0^T, \ldots, x_0^T], \quad \tilde{x}_0^\delta \in \mathbb{O},
\]
such that \(\tilde{x}_0 = \tilde{x}_0^\delta + \tilde{x}_0^\delta \) and
\[
\begin{bmatrix}
0_{[0,L-1]} \\
\tilde{u}_{[0,L-1]}
\end{bmatrix} = \begin{bmatrix} 0 & I & T_L \end{bmatrix} \begin{bmatrix} \tilde{x}_0^\delta & \tilde{x}_0^\delta \end{bmatrix}
\]
where
\[
T_L = \begin{bmatrix}
D & 0 & 0 & \cdots & 0 \\
CB & D & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
CA^{L-3} & CA^{L-1} & \cdots & D
\end{bmatrix},
\]
\[
O_L = \begin{bmatrix} C^\top A_C & (CA_C)^\top & \cdots & (CA_C)^{(L-1)} \end{bmatrix}^\top.
\]

Further, using the Cayley-Hamilton theorem one can show that \(O \subseteq \ker O_L\) for all \(L \in \mathbb{N}_+\). Hence \((27)\) implies
\[
O_L x_0 = 0_{Lp}.
\]
Since \(x_0^\delta \in R + K[x_0^T, x_0^T, \ldots, x_0^T]\), the first statement implies that there exists \(g \in \mathbb{R}^{\sum_{i=1}^{T-1}(T'-L+1)}\) such that
\[
\begin{bmatrix}
\tilde{x}_0^\delta \\
\tilde{u}_{[0,L-1]}
\end{bmatrix} = \begin{bmatrix} H_1(x_{0,T'-L}) & \cdots & H_1(x_{0,T'-L}) \\
H_L(u_{0,T'-1}) & \cdots & H_L(u_{0,T'-1}) \end{bmatrix} g.
\]
Notice that
\[
\begin{bmatrix}
0 & I \\
O_L & T_L
\end{bmatrix}
\begin{bmatrix}
H_1(u_1^{i,0,T'-1}) \\
H_L(u^{i,0,T'-1})
\end{bmatrix} = 
\begin{bmatrix}
H_1(u_1^{i,0,T'-1}) \\
H_L(y_{0,T'-1})
\end{bmatrix}.
\]
for \(i = 1, \ldots, \tau\). Substituting (30), (31) and (32) into (28) gives \(4\), thus completing the proof.

*Proof of Corollary 2.* The first statement of Theorem 1 implies that, for any \(z \in \mathcal{R} + \mathcal{K}[x_0, x_0, \ldots, x_0] \) and \(v \in \mathbb{R}^m\), there exists \(g \in \mathbb{R}^{\sum_{i=1}^n T_i} \) such that \(\begin{bmatrix}
z \\
v
\end{bmatrix} = Sg\). Using the definition of matrix \(S\) and \(X\) we can show
\[
\begin{bmatrix}
A \\
B
\end{bmatrix}
\begin{bmatrix}
z \\
v
\end{bmatrix} = [A \ B] Sg = Xg.
\]
Further, notice that
\[
\begin{bmatrix}
\hat{A} \\
\hat{B}
\end{bmatrix}
\begin{bmatrix}
z \\
v
\end{bmatrix} = XS^T Sg = Xg = \begin{bmatrix}
A \\
B
\end{bmatrix}
\begin{bmatrix}
z \\
v
\end{bmatrix}.
\]
This completes the proof by considering the following two cases where: 1) \(z = 0\) and \(v\) ranges over \(\mathbb{R}^m\), and 2) \(v = 0\) and \(z\) ranges over \(\mathcal{R} + \mathcal{K}[x_0, x_0, \ldots, x_0]\).

Markov parameters computation in Section III-B. Let \(\{u_i^{0,T'-1}, y_i^{0,T'-1}\}_{i=1}^{\tau}\) be input-output trajectories of the system described by (14), such that inputs \(\{u_i^{0,T'-1}\}_{i=1}^{\tau}\) are collectively persistently exciting of order \((N + 1)\bar{n} + 1\). Let \(n = N\bar{n}, m = N\bar{m}, p = M\bar{n}\). Since \((\hat{A}, \hat{B})\) in (14) is controllable, one can verify that \(\mathcal{R} + \mathcal{K}[x_0, x_0, \ldots, x_0] = \mathbb{R}^m\), regardless of the values of \(x_0, x_0, \ldots, x_0\). Using Corollary 2 we know that there exists a matrix \(G_k \in \mathbb{R}^{\sum_{i=1}^n (T_i-n)} \times m\) such that,
\[
\begin{bmatrix}
H_{n+1}(u_1^{0,T'-1}) \\
\vdots \\
H_{n+1}(u_0^{0,T'-1}) \\
H_{n+1}(y_1^{0,T'-1}) \\
H_{n+1}(y_0^{0,T'-1})
\end{bmatrix}G_k
= \begin{bmatrix}
0_{m(n-k) \times m} \\
I_m \\
0_{(p+n-km-kp) \times m} \\
M_0 \\
\vdots \\
M_k^{T}
\end{bmatrix}
\]
(33)
for all \(k = 0, 1, \ldots, \bar{n} + 1\), where \(M_0 = 0_{p \times m}\) and \(M_k\) is given by (15); also see [7, Sec. 4.5] for details. Next, given \(M_0, \ldots, M_{k-1}\), we can compute \(M_k\) by first solving the first \(m(n + 1) + pm\) equations in (33) for matrix \(G_k\), then substituting the solution into the last \(p\) equations in (33). By repeating this process for \(k = 1, \ldots, \bar{n} + 1\) we obtain Markov parameters (15). Using Kalman decomposition one can verify that Markov parameters of system (14) are the same as those of a reduced order controllable and observable system with state dimension less than \(n\). Hence matrix \(M_k\) obtained this way is unique; see [7, Prop. 1].

REFERENCES

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

[2] 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 Syst. Lett., vol. 4, no. 3, pp. 602–607, 2020.

[3] I. Markovsky, J. C. Willems, P. Rapisarda, and B. L. De Moor, “Algorithms for deterministic balanced subspace identification,” Automatica, vol. 41, no. 5, pp. 755–766, 2005.

[4] T. Katayama, Subspace methods for system identification. Springer Science & Business Media, 2006.

[5] I. Markovsky, “From data to models,” in Low-Rank Approximation. Springer, 2019, pp. 37–70.

[6] 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.

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

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

[9] A. Bisoffi, C. De Persis, and P. Tesi, “Data-based guarantees of set invariance properties,” arXiv preprint arXiv:1911.12293 [eecs.SY], 2019.

[10] 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.

[11] L. Huang, J. Coulson, J. Lygeros, and F. Dörfler, “Data-enabled predictive control for grid-connected power converters,” in IEEE Conf. Decision Control. IEEE, 2019, pp. 8130–8135.

[12] J. Coulson, J. Lygeros, and F. Dörfler, “Data-enabled predictive control: In the shallows of the deep,” in Eur. Control Conf. IEEE, 2019, pp. 307–312.

[13] 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.

[14] J. Berberich, J. Koehler, M. A. Muller, and F. Allgöwer, “Data-driven model predictive control with stability and robustness guarantees,” IEEE Transactions on Automatic Control, pp. 1–1, 2020.

[15] A. Alilihouy and J. Cortés, “Data-based receding horizon control of linear network systems,” arXiv preprint arXiv:2003.09813, 2020.

[16] M. Yin, A. Iannelli, and R. S. Smith, “Maximum likelihood estimation in data-driven modeling and control,” arXiv preprint arXiv:2011.00925 [eess.SY], 2020.

[17] F. Fabiani and P. J. Goulart, “The optimal transport paradigm enables data compression in data-driven robust control,” arXiv preprint arXiv:2005.09393 [eess.SY], 2020.

[18] V. K. Mishra, I. Markovsky, and B. Grossmann, “Data-driven tests for controllability,” IEEE Control Systems Lett., vol. 5, no. 2, pp. 517–522, 2020.

[19] I. Markovsky and F. Dörfler, “Identifiability in the behavioral setting,” Dept. ELEC, Vrije Universiteit Brussel, Tech. Rep., 2020.

[20] R. A. Horn and C. R. Johnson, Matrix analysis. Cambridge university press, 2012.

[21] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert, “Constrained model predictive control: Stability and optimality,” Automatica, vol. 36, no. 6, pp. 789–814, 2000.

[22] D. Q. Mayne, “Model predictive control: Recent developments and future promise,” Automatica, vol. 50, no. 12, pp. 2967–2986, 2014.

[23] H. J. Van Waarde, J. Eising, H. L. Trentelman, and M. K. Camlibel, “Data informativity: a new perspective on data-driven analysis and control,” IEEE Trans. Autom. Control, 2020.

[24] G. Wen, Z. Duan, W. Ren, and G. Chen, “Distributed consensus of multi-agent systems with general linear node dynamics and intermittent communications,” Int. J. Robust. Nonlinear Control, vol. 24, no. 16, pp. 2438–2457, 2014.

[25] M. Verhaegen and V. Verdult, Filtering and system identification: a least squares approach. Cambridge University Press, 2007.