Data-Driven Criteria for Detectability and Observer Design for LTI Systems

Mishra, Vikas Kumar; Van Waarde, Henk J.; Bajcinca, Naim

Published in:
2022 IEEE 61st Conference on Decision and Control, CDC 2022

DOI:
10.1109/CDC51059.2022.9992632

2022 IEEE 61st Conference on Decision and Control (CDC)
December 6-9, 2022. Cancún, Mexico

Data-Driven Criteria for Detectability and Observer Design for LTI Systems

Vikas Kumar Mishra[1], Henk J. van Waarde[2], and Naim Bajcinca[1]

2022 IEEE 61st Conference on Decision and Control (CDC) | 978-1-6654-6761-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/CDC51059.2022.9992632

***Abstract*— We study the problems of determining the detectability and designing a state observer for linear time-invariant systems from measured data. First, we establish algebraic criteria to verify the detectability of the system from noise-free data. Then, we formulate data-driven linear matrix inequality-based conditions for observer design. Finally, we give conditions to infer the detectability of the system from noisy data.**

I. Introduction

It is known that detectability is necessary and sufficient for observer design of linear time-invariant (LTI) systems. An observer is an algorithm that estimates the state variables from input/output measurements of a known system and the process of building an observer is called observer design. In this paper, we are interested in developing data-driven criteria to verify the observability/detectability of an LTI system and to design a state observer for the system.

The notion of observers dates back to the seminal contribution by Luenberger [1]. Since then the theory of observers has witnessed tremendous developments in several directions [2], [3]. Recently, due to the ever-increasing interest in machine learning tools across communities, control theory has regained a considerable amount of interest in data-driven control. Data-driven control essentially means data-driven analysis or designing control laws directly from the measured data, without explicit model identification (estimating model parameters from measured data). The root of this surge in data-driven control may be traced back to the idea of viewing systems as a set of "valid" trajectories. Here, valid refers to the fact that the trajectories must satisfy a set of dynamical equations that defines the system. This idea first conceived by Willems in the 1980s is known as the *behavioral system theory*.

About almost two decades ago, Willems and co-workers proved a result that provides sufficient conditions under which the set of input/output trajectories of a deterministic linear time-invariant (LTI) system can be recovered from a single measured input/output trajectory [4, Theorem 1]. This result is popularly known as the *fundamental lemma*. The fundamental lemma is a cornerstone of data-driven control theory and has been further extended in many directions such as the development of necessary and sufficient conditions [5], [6], uncontrollable LTI systems [7], [8], stochastic LTI systems [9], multiple trajectories [10], several classes of nonlinear systems [11]–[13] to name a few. Further, the fundamental lemma has been directly used in tackling several problems such as data-driven simulation and control [14], data-driven stabilization [15], data-driven predictive control [16]–[18], data-driven observers [19], and data-driven input design [20] to name a few. See a more recent survey [21] for an overview.

However, here we are interested in an alternative approach, namely *data informativity*. Inspired by robust control, in this approach, instead of recovering/investigating the true system, we investigate the set of systems that explain the given data and then tackle the control problem at hand for this set of systems, and thus the problem is naturally solved for the true system that generated the data [22]. This approach has been used in solving several control problems such as verifying structural properties of LTI systems [23], [24], closed-loop parametrization from open-loop data [25], stabilization problem [26], dissipativity problem [27], and algebraic regulator problem [28] to name a few.

In this paper, we develop data-driven criteria to determine the observability/detectability and to design a state observer for LTI systems. A work similar to our observability result has been developed in [24], wherein, unlike ours, it is assumed that the output matrix is known. Data-driven observers have been considered in [19]. However, unlike ours, their approach is based on the fundamental lemma. Further, we will provide necessary and sufficient conditions to verify the detectability of the system from noisy data.

Our notation is standard. The set of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. 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 any matrix $A \in \mathbb{R}^{k\times m}$ are denoted by $A^\top$ and $A^\dagger$, respectively. The identity and zero matrices of appropriate dimensions are denoted by I and 0, respectively. For any symmetric matrix $A \in \mathbb{R}^{k\times k}$, $\mathrm{In}(A)$ denotes its *inertia* and is defined as $\mathrm{In}(A) = (\sigma_-, \sigma_0, \sigma_+)$, where σ_-, σ_0, and σ_+ denotes, respectively, the number of negative, zero, and positive eigenvalues of the matrix A. If A is positive definite or positive semidefinite, we denote it by $A > 0$ or $A \geq 0$. The symbol $\Lambda(A, B) := \{\lambda \in \mathbb{C} \mid \det(A - \lambda B) = 0\}$ denotes the set of generalized eigenvalues of a pair of square matrices (A, B). Finally, $\mathrm{im}(A)$ denotes the image of the matrix A.

The rest of the paper is organized as follows. We study

This work was supported by the German Federation of Industrial Research Associations (AiF) within the scope of the projects Virman with the grant number KK5048411LF1.

[1]The authors are with the Department of Mechanical and Process Engineering, Technische Universität Kaiserlautern, Gottileb-Daimler-Straße 42, 67663 Kaiserslautern, Germany (email: {vikas.mishra, naim.bajcinca}@mv.uni-kl.de).

[2]Henk J. van Waarde is with the Bernoulli Institute for Mathematics, Computer Science, and Artificial Intelligence at the University of Groningen, Netherlands (email: h.j.van.waarde@rug.nl).

observability/detectability from noise-free data in Section III, and then observer design in Section IV. In Section V, we investigate detectability from noisy data. Section VI contains our numerical experiments. Finally, Section VII contains concluding remarks and future research directions.

II. Problem Formulation

This paper is devoted to study the problems of deriving data-driven criteria for observability/detectability and observer design for the linear time-invariant (LTI) system

$$x(t+1) = A_s x(t) \tag{1a}$$

$$y(t) = C_s x(t). \tag{1b}$$

Here, $x(t) \in \mathbb{R}^n$ is the state vector and $y(t) \in \mathbb{R}^p$ the output vector. Parameter matrices A_s, C_s are of appropriate dimensions. In the following, we recall the Hautus lemma for observability/detectability of system (1).

Lemma 1: System (1) is observable if and only if for all $\lambda \in \mathbb{C}$

$$\operatorname{rank} \begin{bmatrix} A_s - \lambda I \\ C_s \end{bmatrix} = n \tag{2}$$

holds. Likewise, the system is detectable if and only if (2) holds for all $\lambda \in \mathbb{C}$ with $|\lambda| \geq 1$.

A full-order observer for system (1) is described by

$$\hat{x}(t+1) = A_s \hat{x}(t) + L(y(t) - \hat{y}(t)) \tag{3a}$$

$$\hat{y}(t) = C_s \hat{x}(t). \tag{3b}$$

Here, $\hat{x}(t) \in \mathbb{R}^n$ is an estimate of $x(t)$ and L is a to-be-designed matrix of appropriate dimensions. System (3) is called an asymptotic observer for (1) if the error vector $e(t) = x(t) - \hat{x}(t) \to 0$ as $t \to \infty$. Using (1) and (3), the error dynamics is given by

$$e(t+1) = (A_s - LC_s)e(t) \tag{4}$$

Thus, we aim at designing L such that $A_s - LC_s$ is Schur, *i.e.*, all the eigenvalues of $A_s - LC_s$ lie inside the unit disk. We now recall the following standard result that provides necessary and sufficient conditions for observer design.

Theorem 1: [2] The following statements are equivalent:

1) There exists asymptotic observer (3) for system (1);
2) Matrix pair (A_s, C_s) is detectable;
3) There exists L such that $A_s - LC_s$ is Schur;
4) There exist L and $P = P^\top > 0$ such that

$$P - (A_s - LC_s)^\top P (A_s - LC_s) > 0.$$

Because we are interested in data-driven analysis and design problems for system (1), we assume that the true parameter matrices A_s and C_s are unknown. However, we have access to the state/output data. Following the notation of [22], system (1) can be rewritten as

$$X_+ = A_s X_- \tag{5a}$$

$$Y_- = C_s X_-, \tag{5b}$$

with

$$\begin{aligned} X &:= \begin{bmatrix} x(0) & x(1) & \dots & x(T-1) & x(T) \end{bmatrix} \in \mathbb{R}^{n \times (T+1)} \\ X_+ &:= \begin{bmatrix} x(1) & x(2) & \dots & x(T-1) & x(T) \end{bmatrix} \in \mathbb{R}^{n \times T} \\ X_- &:= \begin{bmatrix} x(0) & x(1) & x(2) & \dots & x(T-1) \end{bmatrix} \in \mathbb{R}^{n \times T} \\ Y_- &:= \begin{bmatrix} y(0) & y(1) & y(2) & \dots & y(T-1) \end{bmatrix} \in \mathbb{R}^{p \times T}. \end{aligned}$$

Naturally, the set of systems explaining the data is given by

$$\Sigma = \{(A, C) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{p \times n} \mid \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} X_- \}. \tag{6}$$

Definition 1: We say that the data (X, Y_-) are *informative for observability/detectability* if every pair $(A, C) \in \Sigma$ is observable/detectable.

The system is *identifiable* if one can uniquely recover the parameters (A, C) from the data. The following proposition on identifiability of system (1) is a straightforward consequence of [22, Proposition 6].

Proposition 1: System (1) is identifiable if and only if X_- is of full row rank. In this case, we have

$$A_s = X_+ X_-^\dagger \text{ and } C_s = Y_- X_-^\dagger. \tag{7}$$

Remark 1: The observer design discussed in this paper relies on the state data X, which are usually obtained via an observer. However, these data are collected offline and once the observer is constructed, it can be used to estimate the state at any point in time and also in online experiments. See also [19, Remark 4]. Further, this work can be seen as a stepping stone to a more general problem, where only (noisy) input/output data will be used to design an observer.

We are now in a position to state the problems that we tackle in this paper.

1) Find conditions under which the noise-free data (X, Y_-) are informative for observability/detectability.
2) If the noise-free data (X, Y_-) are informative for detectability, design an asymptotic observer for (1) on the basis of the given data.
3) Extend 1) to noisy data, as explained in more detail in Section V.

III. Data-Driven Tests for Observability/Detectability

In this section, we develop a data-driven tests for observability/detectability of system (1). First we show that the full row of X_- is a necessary condition for detectability and then give data-driven necessary and sufficient conditions for observability/detectability of the system.

Theorem 2: If the data (X, Y_-) are informative for detectability, then X_- is of full row rank.

Prior to proving this theorem, we first state and prove an auxiliary lemma.

Lemma 2: Let $A \in \mathbb{R}^{n \times n}$ and nonzero $\xi \in \mathbb{R}^n$. Then, for every $\delta > 0$ there exists $\beta \in \mathbb{R}$ such that $\rho(A + \beta \xi \xi^\top) > \delta$, where $\rho(\cdot)$ denotes spectral radius.

Proof: Assume on contrary that there exists $\delta > 0$ such that

$$\rho(A + \beta \xi \xi^\top) \leq \delta \quad \text{for every } \beta \in \mathbb{R}.$$

This implies

$$\rho(\frac{1}{\beta}A + \xi\xi^\top) \leq \frac{\delta}{\beta} \quad \text{for every } \beta > 0.$$

Now, taking the limit $\beta \to \infty$ and using the continuity of the spectral radius, we can conclude that $\rho(\xi\xi^\top) = 0$, which is a contradiction to the fact that $\xi \neq 0$. Hence, lemma holds. ∎

Proof: [Proof of Theorem 2] Assume on the contrary that $\xi^\top X_- = 0$ for $\xi \neq 0$. Let $(A, C) \in \Sigma$. By Lemma 2, we can choose a $\beta \in \mathbb{R}$ such that

$$\rho(A + \beta\xi\xi^\top) \geq 1 \quad \text{and} \tag{8}$$
$$\rho(A + \beta\xi\xi^\top) > \rho(A) \tag{9}$$

Let $\eta \in \mathbb{C}^n$ be such that

$$(A + \beta\xi\xi^\top)\eta = \lambda\eta, \tag{10}$$

where λ is an eigenvalue of $A + \beta\xi\xi^\top$ whose modulus coincides with $\rho(A + \beta\xi\xi^\top)$. Note that $\xi^\top\eta \neq 0$ by (9). Now, we consider three cases:

1) Let $\lambda \in \mathbb{R}$. Then, without loss of generality, $\eta \in \mathbb{R}^n$. Now define
$$\begin{bmatrix} \bar{A} \\ \bar{C} \end{bmatrix} := \begin{bmatrix} A + \beta\xi\xi^\top \\ C - \frac{1}{\xi^\top\eta}C\eta\xi^\top \end{bmatrix}.$$
Then, $(\bar{A}, \bar{C}) \in \Sigma$ and $\begin{bmatrix} \bar{A} - \lambda I \\ \bar{C} \end{bmatrix}\eta = 0$. Thus, the pair $(\bar{A}, \bar{C}) \in \Sigma$ and is not detectable.
2) Let $\lambda \notin \mathbb{R}$, and $\eta = \eta_1 + i\eta_2$, where $\eta_1 \in \mathbb{R}^n$ and $\eta_2 \in \mathbb{R}^n$ are linearly dependent. Then $\eta = (a + ib)\zeta$ with $\zeta \in \mathbb{R}^n$ and $(A + \beta\xi\xi^\top - \lambda I)\zeta = 0$, which implies λ is real and we are back in case 1).
3) Let $\lambda \notin \mathbb{R}$, and $\eta = \eta_1 + i\eta_2$, where $\eta_1 \in \mathbb{R}^n$ and $\eta_2 \in \mathbb{R}^n$ are linearly independent. Since $\xi^\top\eta \neq 0$, we have $\eta \notin \operatorname{im}(X_-)$. We claim that it is without loss of generality to assume that both $\eta_1, \eta_2 \notin \operatorname{im}(X_-)$. For otherwise, we have two cases: either $\eta_1 \in \operatorname{im}(X_-)$, $\eta_2 \notin \operatorname{im}(X_-)$ or $\eta_1 \notin \operatorname{im}(X_-)$, $\eta_2 \in \operatorname{im}(X_-)$. Define
$$\bar{\eta} := (1 + i)\eta = (\eta_1 - \eta_2) + i(\eta_1 + \eta_2).$$
Note that both $(\eta_1 - \eta_2), (\eta_1 + \eta_2) \notin \operatorname{im}(X_-)$. Our claim will be established provided we prove that $(\eta_1 - \eta_2), (\eta_1 + \eta_2)$ are linearly independent, which follows from our assumption that η_1, η_2 are linearly independent. Thus, it is without loss of generality to assume that both $\eta_1, \eta_2 \notin \operatorname{im}(X_-)$. Hence, there exists matrix C_0 such that
$$C_0X_- = 0, \quad C_0\eta_1 = -\operatorname{Re}(C\eta), \quad C_0\eta_2 = -\operatorname{Im}(C\eta).$$
Now define
$$\begin{bmatrix} \bar{A} \\ \bar{C} \end{bmatrix} := \begin{bmatrix} A + \beta\xi\xi^\top \\ C + C_0 \end{bmatrix}.$$
It can be seen that
$$(\bar{A}, \bar{C}) \in \Sigma \quad \text{and} \quad \begin{bmatrix} \bar{A} - \lambda I \\ \bar{C} \end{bmatrix}\eta = 0.$$

Therefore, if X_- is not of full row rank, there exists a pair $(\bar{A}, \bar{C}) \in \Sigma$, which is not detectable. This completes the proof of the theorem. ∎

The following theorem provides necessary and sufficient conditions for observability/detectability.

Theorem 3: The data (X, Y_-) are informative for observability if and only if for all $\lambda \in \mathbb{C}$

$$\operatorname{rank} \begin{bmatrix} X_+ - \lambda X_- \\ Y_- \end{bmatrix} = n \tag{11}$$

holds. Likewise, the data are informative for detectability if and only if (11) holds for all $\lambda \in \mathbb{C}$ with $|\lambda| \geq 1$.

Proof: We will prove the result for observability. The result for detectability can be proven essentially in a similar fashion. To prove the "if" part, note that (11) implies

$$\operatorname{rank} \begin{bmatrix} A - \lambda I \\ C \end{bmatrix} X_- = n \quad \forall \lambda \in \mathbb{C} \tag{12}$$

for all $(A, C) \in \Sigma$. The last equality implies

$$\operatorname{rank} \begin{bmatrix} A - \lambda I \\ C \end{bmatrix} \geq n \quad \forall \lambda \in \mathbb{C}. \tag{13}$$

Because $\begin{bmatrix} A - \lambda I \\ C \end{bmatrix}$ has n number of columns, the result follows.

Conversely, suppose that the data (X, Y_-) are informative for observability, but that (11) does not hold. That is, there exists $\lambda \in \mathbb{C}$ such that

$$\operatorname{rank} \begin{bmatrix} X_+ - \lambda X_- \\ Y_- \end{bmatrix} \neq n. \tag{14}$$

This implies that for this λ

$$\operatorname{rank} \begin{bmatrix} A - \lambda I \\ C \end{bmatrix} X_- \neq n. \tag{15}$$

Because X_- is of full row rank (Theorem 2), the above implies

$$\operatorname{rank} \begin{bmatrix} A - \lambda I \\ C \end{bmatrix} < n. \tag{16}$$

That is, the system is not observable. This completes the proof of the theorem. ∎

Remark that to verify the model-based observability using Lemma 1, it is sufficient to test (2) for all λ that are the eigenvalues of the matrix A_s. However, the matrix $\begin{bmatrix} X_+ - \lambda X_- \end{bmatrix}$ in (11) is in general not square. Hence, this idea of eigenvalues cannot be directly used to verify (11). Nevertheless, following [7, Algorithm 1], a similar algorithm can be developed to verify the data-driven observability condition (11).

Algorithm 1: Data-driven observability test

Input: Observed data matrices X and Y_-.

Output: The system is observable/unobservable.

1: Construct the data matrices X_-, X_+ from X.
2: If rank $X_- < n$, the data are not informative for observability.
3: If rank $X_- = n$, let $U^\top X_- V = \begin{bmatrix} S & 0 \end{bmatrix}$ (SVD of X_-).

4: Partition X_+ as $U^\top X_+ V = \begin{bmatrix} X_{+,11} & X_{+,12} \end{bmatrix}$, where $X_{+,11} \in \mathbb{R}^{n\times n}$ and partition Y_- conformably as $Y_- V = \begin{bmatrix} Y_{-,11} & Y_{-,12} \end{bmatrix}$.
5: Compute the generalized eigenvalues of $(X_{+,11}, S)$.
6: Compute the rank of $Y_\lambda := \begin{bmatrix} X_{+,11} - \lambda S & X_{+,12} \\ Y_{-,11} & Y_{-,12} \end{bmatrix}$ for all $\lambda \in \Lambda(X_{+,11}, S)$.
7: If $\operatorname{rank} Y_\lambda = n$ for all $\lambda \in \Lambda(X_{+,11}, S)$, the data are informative for observability.

It is important to note that we can compute the rank of the matrix Y_λ for all $\lambda \in \Lambda(X_{+,11}, S)$ with $|\lambda| \geq 1$ to check whether or not the system is detectable.

IV. Data-Driven Observer Design

In this section, first, we give another condition under which the noise-free data (X, Y_-) is informative for detectability. Then, by using this result, we build a data-driven asymptotic state observer.

Theorem 4: The data (X, Y_-) are informative for detectability if and only if X_- is of full row rank and there exist K and $P = P^\top > 0$ such that

$$\begin{bmatrix} P & (PX_+X_-^\dagger - KY_-X_-^\dagger)^\top \\ PX_+X_-^\dagger - KY_-X_-^\dagger & P \end{bmatrix} > 0. \tag{17}$$

Proof: "if" part: Suppose that X_- is of full row rank and (17) is feasible. From Schur complement lemma, (17) is equivalent to

$$P - (PX_+X_-^\dagger - KY_-X_-^\dagger)^\top P^{-1}(PX_+X_-^\dagger - KY_-X_-^\dagger) > 0.$$

Substituting $L = P^{-1}K$, equivalently $K = PL$, the above inequality reduces to

$$P - (X_+X_-^\dagger - LY_-X_-^\dagger)^\top P (X_+X_-^\dagger - LY_-X_-^\dagger) > 0. \tag{18}$$

Because X_- is of full row rank, from Proposition 1, $\Sigma = \{(A_s, C_s)\}$ and A_s, C_s are given by (7). In view of Theorem 1, together (18) and (7) imply that the data (X, Y_-) are informative for detectability.

"Only if" part: Suppose that the data (X, Y_-) are informative for detectability. Then, from Theorem 2, X_- is of full row rank. Now, from Proposition 1 and Theorem 1, it follows that there exist L and $P = P^\top > 0$ such that (18) holds. Substituting $K = PL$, equivalently $L = P^{-1}K$ and by using Schur complement lemma, it can be shown that (18) is equivalent to (17). Thus, (17) is feasible. ■

In view of Theorem 4, we can design a full-order data-driven observer as follows

$$\hat{x}(t+1) = X_+X_-^\dagger \hat{x}(t) + L(y(t) - \hat{y}(t)) \tag{19a}$$
$$\hat{y}(t) = Y_-X_-^\dagger \hat{x}(t), \tag{19b}$$

where $L = P^{-1}K$ can be obtained by solving LMI (17) for P and K.

Remark 2: Although Theorem 4 provides a way to compute the observer gain matrix L from the data, this approach to designing an observer is equivalent to the model-based approach. The reason is that here we need X_- to be of full row rank, which is equivalent to the identifiability of the system (*cf.* Proposition 1).

V. Detectability From Noisy Data

Consider the system

$$x(t+1) = A_s x(t) + w(t) \tag{20a}$$
$$y(t) = C_s x(t) + v(t). \tag{20b}$$

Here, $w(t) \in \mathbb{R}^n$ is the process noise and $v(t) \in \mathbb{R}^p$ is the measurement noise. Following our notation (see Section II), system (20) can be rewritten as

$$X_+ = AX_- + W_- \tag{21a}$$
$$Y_- = CX_- + V_-, \tag{21b}$$

where the noise matrices W_- and V_- are defined as

$$W_- := \begin{bmatrix} w(0) & w(1) & \dots & w(T-1) \end{bmatrix} \in \mathbb{R}^{n\times T}$$
$$V_- := \begin{bmatrix} v(0) & v(1) & \dots & v(T-1) \end{bmatrix} \in \mathbb{R}^{p\times T}.$$

We assume that W_- and V_- satisfy

$$\begin{bmatrix} I \\ \hline \begin{bmatrix} W_-^\top & V_-^\top \end{bmatrix} \end{bmatrix}^\top \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \begin{bmatrix} I \\ \hline \begin{bmatrix} W_-^\top & V_-^\top \end{bmatrix} \end{bmatrix} \geq 0, \tag{22}$$

where $\Phi_{ij},\ i,j = 1,2$, are known matrices of appropriate dimensions with $\Phi_{11} = \Phi_{11}^\top$, $\Phi_{22} = \Phi_{22}^\top < 0$, and $\Phi_{11} - \Phi_{12}\Phi_{22}^{-1}\Phi_{12}^\top > 0$. For a discussion on this assumption, we refer the readers to [26], [27]. In view of (21), inequality (22) can be reformulated as inequality (23). Thus, under the noise model (22), the set of systems given by matrix pair (A, C) that explain the data is given by

$$\widetilde{\Sigma} = \left\{ (A, C) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{p\times n} \mid \text{(23) is satisfied} \right\}. \tag{24}$$

Definition 2: We say that the data (X, Y_-) are *informative for quadratic detectability* if there exist matrices L and $P = P^\top > 0$ such that

$$P - (A - LC)^\top P (A - LC) > 0 \tag{25}$$

for all $(A, C) \in \widetilde{\Sigma}$.

Inequality (25) can be written as

$$\begin{bmatrix} I \\ \hline A \\ C \end{bmatrix}^\top \left[\begin{array}{c|c} P & 0 \\ \hline 0 & -\begin{bmatrix} I \\ -L^\top \end{bmatrix} P \begin{bmatrix} I \\ -L^\top \end{bmatrix}^\top \end{array} \right] \begin{bmatrix} I \\ \hline A \\ C \end{bmatrix} > 0. \tag{26}$$

Thus, to find conditions for data informativity for quadratic detectability is equivalent to finding conditions under which there exist L and $P = P^\top > 0$ such that inequality (26) holds for all matrix pairs (A, C) with the satisfaction of inequality (23). Our idea is to use the matrix S-lemma to tackle this problem. To this end, we recall the matrix S-lemma [26, Section III].

Lemma 3 (Matrix S-lemma): Let two symmetric matrices $M, N \in \mathbb{R}^{(k+m)\times(k+m)}$ be given such that

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^\top & M_{22} \end{bmatrix}, \ N = \begin{bmatrix} N_{11} & N_{12} \\ N_{12}^\top & N_{22} \end{bmatrix}.$$

$$\left[\begin{array}{c} I \\ \hline \begin{bmatrix} A^\top & C^\top \end{bmatrix} \end{array}\right]^\top \left[\begin{array}{c|c} I & \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} \\ \hline 0 & -X_- \end{array}\right] \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \left[\begin{array}{c|c} I & \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} \\ \hline 0 & -X_- \end{array}\right]^\top \left[\begin{array}{c} I \\ \hline \begin{bmatrix} A^\top & C^\top \end{bmatrix} \end{array}\right] \geq 0. \tag{23}$$

Suppose that $M_{22} \leq 0,\ N_{22} \leq 0$, and $\ker N_{22} \subseteq \ker N_{12}$. Further, assume that there exists matrix $\bar{Z} \in \mathbb{R}^{m \times k}$ that satisfies the Slater condition

$$\begin{bmatrix} I \\ \bar{Z} \end{bmatrix}^\top N \begin{bmatrix} I \\ \bar{Z} \end{bmatrix} > 0. \tag{27}$$

Then

$$\begin{bmatrix} I \\ Z \end{bmatrix}^\top M \begin{bmatrix} I \\ Z \end{bmatrix} > 0 \ \forall\ Z \in \mathbb{R}^{m \times k} \text{ with } \begin{bmatrix} I \\ Z \end{bmatrix}^\top N \begin{bmatrix} I \\ Z \end{bmatrix} \geq 0$$

if and only if there exist scalars $\alpha \geq 0$ and $\beta > 0$ such that

$$M - \alpha N \geq \begin{bmatrix} \beta I & 0 \\ 0 & 0 \end{bmatrix}.$$

We remark that (26) is in terms of (A, C) while (23) is in terms of the transposed matrices $A^\top$ and $C^\top$. We thus need a dualization result to transpose again the matrices $A^\top$ and $C^\top$ in order to use the matrix S-lemma. This can be achieved by the following lemma.

Lemma 4 (Dualization lemma): Define the matrix

$$\Psi = \left[\begin{array}{c|c} I & \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} \\ \hline 0 & -X_- \end{array}\right] \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{bmatrix} \left[\begin{array}{c|c} I & \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} \\ \hline 0 & -X_- \end{array}\right]^\top \tag{28}$$

$$=: \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}, \tag{29}$$

where

$$\begin{aligned} \Psi_{11} &= \Phi_{11} + \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} \Phi_{12}^\top + \left(\Phi_{12} + \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} \Phi_{22} \right) \begin{bmatrix} X_+ \\ Y_- \end{bmatrix}^\top \\ \Psi_{12} &= \left(\Phi_{12} + \begin{bmatrix} X_+ \\ Y_- \end{bmatrix} \Phi_{22} \right) X_-^\top \\ \Psi_{21} &= \Psi_{12}^\top \\ \Psi_{22} &= X_- \Phi_{22} X_-^\top. \end{aligned}$$

Assume that $\text{In}(\Psi) = (n, 0, n+p)$. Given a pair (A, C), the inequality (23) holds if and only if

$$\left[\begin{array}{c} I \\ \hline A \\ C \end{array}\right]^\top N \left[\begin{array}{c} I \\ \hline A \\ C \end{array}\right] \geq 0, \tag{30}$$

with

$$N = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix} \Psi^{-1} \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}. \tag{31}$$

Proof: Note that $\Psi_{11} = \Psi_{11}^\top$ and $\Psi_{22} = \Psi_{22}^\top$. Further, by our assumption the matrix Ψ has the correct inertia required for [27, Lemma 3]. The lemma thus follows from [27, Lemma 3]. ■

Remark that the assumption $\text{In}(\Psi) = (n, 0, n+p)$ is a combined assumption on the noise model as well as the data matrices. Relaxing this assumption is a topic of future research. We define

$$\Psi^{-1} = \begin{bmatrix} \widetilde{\Psi}_{11} & \widetilde{\Psi}_{12} \\ \widetilde{\Psi}_{21} & \widetilde{\Psi}_{22} \end{bmatrix}$$

and thus

$$N = \begin{bmatrix} -\widetilde{\Psi}_{22} & \widetilde{\Psi}_{21} \\ \widetilde{\Psi}_{12} & -\widetilde{\Psi}_{11} \end{bmatrix}, \tag{32}$$

where

$$\begin{aligned} \widetilde{\Psi}_{11} &= (\Psi/\Psi_{22})^{-1} \\ \widetilde{\Psi}_{12} &= -(\Psi/\Psi_{22})^{-1} \Psi_{12} \Psi_{22}^{-1} \\ \widetilde{\Psi}_{21} &= -\Psi_{22}^{-1} \Psi_{12}^\top (\Psi/\Psi_{22})^{-1} \\ \widetilde{\Psi}_{22} &= \Psi_{22}^{-1} + \Psi_{22}^{-1} \Psi_{12}^\top (\Psi/\Psi_{22})^{-1} \Psi_{12} \Psi_{22}^{-1}. \end{aligned} \tag{33}$$

Here $\Psi/\Psi_{22} = \Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{12}^\top$ is the Schur complement of Ψ_{22}. In view of the symbols used in the matrix S-lemma, we define

$$M = \left[\begin{array}{c|c} P & 0 \\ \hline 0 & -\begin{bmatrix} I \\ -L^\top \end{bmatrix} P \begin{bmatrix} I \\ -L^\top \end{bmatrix}^\top \end{array}\right] \tag{34}$$

and N as in Eq. (32).

We now verify the assumptions of the matrix S-lemma. Because $P > 0$,

$$M_{22} = -\begin{bmatrix} I \\ -L^\top \end{bmatrix} P \begin{bmatrix} I \\ -L^\top \end{bmatrix}^\top \leq 0.$$

Next, under the inertia assumption on Ψ,

$$N_{22} = -\widetilde{\Psi}_{11} \leq 0.$$

Finally, $\ker N_{22} \subseteq \ker N_{12}$ follows directly from the fact that N_{22} is nonsingular. Combining the above discussion, the following theorem is stated, which gives necessary and sufficient conditions for the data being informative for quadratic detectability.

Theorem 5: Let $\text{In}(\Psi) = (n, 0, n+p)$. Then, the data (X, Y_-) are informative for quadratic detectability if and only if the LMI (35), where $\widetilde{\Psi}_{11}, \widetilde{\Psi}_{12}, \widetilde{\Psi}_{21}, \widetilde{\Psi}_{22}$ are defined in (33), is solvable for matrices $P = P^\top > 0$, K, and scalars $\alpha \geq 0$ and $\beta > 0$.

Proof: Note that $\text{In}(\Psi) = (n, 0, n+p)$ implies $\text{In}(N) = (n+p, 0, n)$, where N is given in (32). Thus, condition (27) holds for the matrix N and for some $\bar{Z} \in \mathbb{R}^{n \times (n+p)}$.

To prove the "if" part, let LMI (35) be solvable for $P = P^\top > 0$, K, and scalars $\alpha \geq 0$, $\beta > 0$. Then, by using $K = PL$, where $L \in \mathbb{R}^{n \times p}$, and the Schur complement lemma, it can be shown that (35) is equivalent to

$$M - \alpha N \geq \begin{bmatrix} \beta I & 0 \\ 0 & 0 \end{bmatrix}, \tag{36}$$

$$\left[\begin{array}{c|ccc} P-\beta I & 0 & 0 & 0 \\ \hline 0 & -P & K & 0 \\ 0 & K^\top & 0 & K^\top \\ 0 & 0 & K & P \end{array}\right] - \alpha \left[\begin{array}{cc} I & 0 \\ \hline 0 & I \\ 0 & 0 \end{array}\right] \begin{bmatrix} -\widetilde{\Psi}_{22} & \widetilde{\Psi}_{21} \\ \widetilde{\Psi}_{12} & -\widetilde{\Psi}_{11} \end{bmatrix} \left[\begin{array}{cc} I & 0 \\ \hline 0 & I \\ 0 & 0 \end{array}\right]^\top \geq 0 \tag{35}$$

where M and N are defined in (34) and (32). Now, from Lemma 3, it follows that (26) holds for all pairs (A, C) satisfying (30). Further, from Lemma 4, (30) holds if and only if (23) holds. Thus, (26) holds for all pairs $(A, C) \in \widetilde{\Sigma}$. That is, the data are informative for quadratic detectability.

Conversely, suppose that the data are informative for quadratic detectability. That is, there exist L and $P = P^\top > 0$ such that (26) holds for all pairs (A, C) with the satisfaction of (23). From Lemma 4, (23) holds if and only if (30) holds. Now, from Lemma 3, there exist scalars $\alpha \geq 0$ and $\beta > 0$ such that (36) holds, where M and N are defined in (34) and (32). By using $L = P^{-1}K$, where $K \in \mathbb{R}^{n\times p}$, and the Schur complement lemma, it follows that (36) is equivalent to (35). This completes the proof. ∎

Unlike in the noise-free data setting, data informativity for quadratic detectability does not imply that $\widetilde{\Sigma}$ is a singleton set. Therefore, although we could characterize quadratic detectability, which is a necessary condition for observer design, for the entire set of systems $\widetilde{\Sigma}$, it is in general not possible to obtain a single observer for all systems in $\widetilde{\Sigma}$. The reason is that the observer dynamics *depend* on the particular system $(A, C) \in \widetilde{\Sigma}$. Designing an observer from noisy data is a subject of future research.

VI. Numerical Experiments

We consider a pendulum, see for instance [3, Example 10.5.4], whose dynamics is given by

$$m\ell^2\ddot{\theta} + m\ell g \sin\theta = 0, \tag{37}$$

where m is the mass of the pendulum, ℓ its length, θ the angle, and g the constant of gravity. For the purposes of simulation, let us take $m = 1$, $\ell = 1$, and $g = 9.8$. For small values of θ, $\sin\theta \approx \theta$. In that case, taking $x_1 = \theta$ and $x_2 = \dot{\theta}$, the nonlinear second order dynamics (37) can be written as the first order linear dynamics:

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -9.8 & 0 \end{bmatrix} x, \tag{38}$$

where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Assuming that we can observe the variable x_1, we can write the output equation as

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} x. \tag{39}$$

We then discretize (38)-(39) by using the built-in function `c2d` of Matlab with sampling time 0.4 s. This yields an unstable discrete-time system

$$x(t+1) = \begin{bmatrix} 0.3132 & 0.3034 \\ -2.973 & 0.3132 \end{bmatrix} x(t) \tag{40a}$$

$$y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t). \tag{40b}$$

We consider Gaussian distributed random initial state with zero mean and unit variance and generate the sampled data of length $T = 20$. It can be verified that system (40) is detectable by using Algorithm 1. Solving (17), we obtain the matrices

$$P = \begin{bmatrix} 2.8369 & -0.0928 \\ -0.0928 & 3.2641 \end{bmatrix} \text{ and } K = \begin{bmatrix} 1.1937 \\ -9.6994 \end{bmatrix}.$$

Hence, by using the backslash operator $\backslash$ of Matlab, we obtain the desired observer gain matrix

$$L = \begin{bmatrix} 0.3238 \\ -2.9624 \end{bmatrix} \text{ from } K = PL,$$

and thus the observer (19) can be designed. The plots for the true and estimated states are given in Fig. 1.

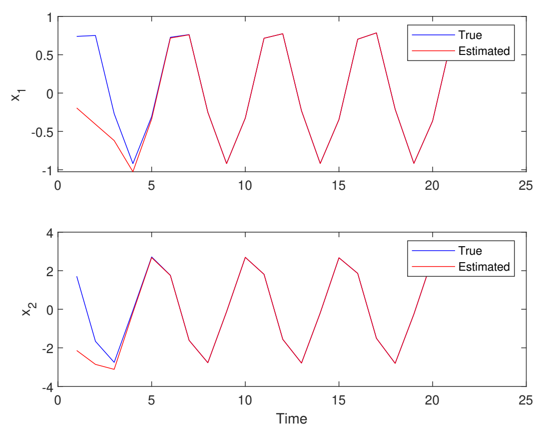

Fig. 1. True and estimated states.

Noisy case

We adapt the set-up of noise-free case above. However, unlike the noise-free case, here we consider system (20), where we assume that the noise samples of both $w(t)$ and $v(t)$ are bounded in norm as

$$\left\| \begin{bmatrix} w^\top(t) & v^\top(t) \end{bmatrix}^\top \right\|_2^2 \leq \kappa \text{ for all } t. \tag{41}$$

It has been shown in [26, Section II] that the above noise bound (41) is satisfied if we take our noise model (22) as $\Phi_{11} = \kappa T I$, $\Phi_{12} = 0$, and $\Phi_{22} = -I$. We choose $\kappa = 0.1$ and use Theorem 5 to infer the quadratic detectability. The matrix Ψ has 2 negative eigenvalues and 3 positive eigenvalues. Thus, the inertia assumption on Ψ is satisfied. Upon solving the LMI (35), we obtain a desired solution as follows.

$$P = \begin{bmatrix} 0.1086 & -0.0045 \\ -0.0045 & 0.0310 \end{bmatrix}, \; K = \begin{bmatrix} 0.0465 \\ -0.0779 \end{bmatrix},$$

$$\alpha = 0.6121,\ \beta = 0.0034,\ L = \begin{bmatrix} 0.3256 \\ -2.4641 \end{bmatrix}.$$

As such, the data are informative for quadratic detectability. We have solved the LMIs in Matlab using Yalmip [29] in both noise-free and noisy cases.

VII. Concluding Remarks and Outlook

We have formulated necessary and sufficient data-driven criteria to verify the observability/detectability of an LTI system. Based on this, we have developed a numerically reliable algorithm to check whether or not the system is observable/detectable. After that, we have considered the problem of data-driven observer design. An LMI-based data-driven method has been developed to design an observer, which can be easily implemented on a computer. Next, we have studied the quadratic detectability from noisy data using a recently developed matrix S-lemma. We have taken the pendulum system as an example to illustrate our developed results.

This work opens up several interesting research questions. So far, we have assumed that we have access to few samples of offline state data; however, it would indeed be interesting to develop analogous results based on the input/output data (*cf.* Remark 1). One can consider the noisy input/output data case as well. Another problem of interest would be to design data-driven reduced-order observers. Further, to generalize these results to different system classes, for instance linear differential-algebraic systems or nonlinear systems would be interesting.

References

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

[2] J. O'Reilly, *Observers for linear systems*. Academic Press, 1983, vol. 170.

[3] J.-W. Polderman and J. Willems, *Introduction to the mathematical theory of systems and control*, 1998, vol. 434.

[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," *Vrije Universiteit Brussel, Tech. Rep*, 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," *arXiv preprint arXiv:2102.02953*, 2021.

[9] G. Pan, R. Ou, and T. Faulwasser, "On a stochastic fundamental lemma and its use for data-driven MPC," *arXiv preprint arXiv:2111.13636*, 2021.

[10] 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.

[11] V. K. Mishra, I. Markovsky, A. Fazzi, and P. Dreesen, "Data-driven simulation for NARX systems," in *2021 29th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1055–1059.

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.

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

[21] 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.

[22] 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*, vol. 65, no. 11, pp. 4753–4768, 2020.

[23] J. Eising and H. L. Trentelman, "Informativity of noisy data for structural properties of linear systems," *Syst. Control Lett.*, vol. 158, p. 105058, 2021.

[24] J. Eising, H. L. Trentelman, and M. K. Camlibel, "Data informativity for observability: An invariance-based approach," in *2020 European Control Conference (ECC)*. IEEE, 2020, pp. 1057–1059.

[25] J. Berberich, A. Koch, C. W. Scherer, and F. Allgöwer, "Robust data-driven state-feedback design," in *Proc. Amer. Control Conf.* IEEE, 2020, pp. 1532–1538.

[26] H. J. van Waarde, M. K. Camlibel, and M. Mesbahi, "From noisy data to feedback controllers: non-conservative design via a matrix S-lemma," *IEEE Trans. Autom. Control*, 2020.

[27] H. J. van Waarde, M. K. Camlibel, P. Rapisarda, and H. L. Trentelman, "Data-driven dissipativity analysis: application of the matrix S-lemma," *arXiv preprint arXiv:2109.02090*, 2021.

[28] H. L. Trentelman, H. J. Van Waarde, and M. K. Camlibel, "An informativity approach to the data-driven algebraic regulator problem," *IEEE Trans. Autom. Control*, 2021.

[29] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in matlab," in *Proc. CACSD Conf.*, Taipei, Taiwan, 2004.