Data-driven Input Reconstruction and Experimental Validation

Jicheng Shi, Yingzhao Lian*, and Colin N. Jones

***Abstract*— This paper proposes a data-driven input reconstruction method from outputs (IRO) based on the Willems' Fundamental Lemma. Given only output measurements, the unknown inputs estimated recursively by the IRO asymptotically converge to the true input without knowing the initial conditions. A recursive IRO and a moving-horizon IRO are developed based respectively on Lyapunov conditions and Luenberger-observer-type feedback, and their asymptotic convergence properties are studied. An experimental study is presented demonstrating the efficacy of the moving-horizon IRO for estimating the occupancy of a building on the EPFL campus via measured carbon dioxide levels.**

I. INTRODUCTION

Input reconstruction estimates unknown inputs based on measured states/outputs, which finds broad application in sensor fault detection, and robust control [1]–[3]. This problem is of particular interest when the real-time/online measurement of inputs is not affordable or is privacy-sensitive. For example, a critical factor in predicting the evolution of the thermal state of a building is the number of occupants, but this value can often not be measured directly by cameras or Wi-Fi due to privacy or cost concerns. Instead, an indirect estimation is commonly deployed based on the measurement of indoor CO2 levels [4]. Another important example is the cutting force of machine tools, whose measurement is only feasible with a dedicated laboratory setup [5].

The input reconstruction problem has been studied in a model-based setup, and various methods have been proposed based on the unknown input observer (UIO) [6], [7], optimal filters [8], the generalized inverse approach [9], sliding mode observers [5] and PI observers [10]. UIO is of special interest in our study, and most methods fall into two categories in the model-based setup. In one class of methods, system states are measured or estimated, and are further used to reconstruct the unknown input by matrix inversion [7] or matrix pencil decomposition [1]. In another category of methods, states and unknown inputs are estimated concurrently, and this estimate can achieve finite step convergence [9].

In practice, the model of the targeted system is usually not available. Instead of running a system identification procedure, the Willems' fundamental lemma offers a direct characterization of the system responses with an informative historical dataset [11]. This characterization provides a convenient interface to data-driven methods, and has been deployed in output prediction [12] and in controller design [13]–[18].

We apply the Willems' fundamental lemma to enable direct input reconstruction with historical I/O data. A similar setup was studied in [19], where the system states are assumed to be measured. Our work removes the requirement of state measurement and achieves unknown input reconstruction directly from output measurements. In order to stress this difference, the approach developed in this paper is termed the input reconstruction method from outputs (IRO), instead of the unknown input observer (UIO). In this paper, we propose two design schemes of a data-driven stable IRO.

In the following, Section II reviews output prediction based on the Willems' Fundamental Lemma, alongside the statement of the IRO problem. The design of a stable data-driven recursive IRO is proposed in Section III, followed by its moving-horizon counterpart in Section IV. The proposed IROs are validated in Section V by simulations and an occupancy estimation experiment in a real-world building, followed by a conclusion in Section VI.

Notation: $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ denotes an identity matrix. $\mathbf{0}_{n,m} \in \mathbb{R}^{n\times m}$ denotes a zero matrix, and the shape information is neglected when there is no confusion. The number of columns and rows of a matrix M are denoted respectively by n_M and m_M such that $M \in \mathbb{R}^{n_M\times m_M}$. Accordingly, $\text{Null}(M)$ denotes its null space. $M^g := \{X|MXM = M\}$ is the set of generalized inverses of matrix M. A strictly positive definite matrix M is denoted by $M \succ 0$. The dimension of a vector s denoted by n_s. Given an ordered sequence of vectors $\{s_t, s_{t+1}, \dots, s_{t+L}\}$, its vectorization is denoted by $s_{t:t+L} = [s_t^\top, \dots, s_{t+L}^\top]^\top$.

II. PRELIMINARIES

Consider a discrete-time linear time-invariant (LTI) system. Its minimal representation, dubbed $\mathcal{B}(A,B,C,D)$, is:

$$x_{t+1} = Ax_t + Bu_t \,, y_t = Cx_t + Du_t \,, \tag{1}$$

whose states, inputs and outputs are denoted by $x \in \mathbb{R}^{n_x}$, $u \in \mathbb{R}^{n_u}$ and $y \in \mathbb{R}^{n_y}$ respectively. Measureable inputs are omitted to simplify the notation. The order of the system is defined as $n(\mathcal{B}(A,B,C,D)) := n_x$. The lag of the system $l(\mathcal{B}(A,B,C,D))$ is defined as the smallest integer ℓ for which the observability matrix $\mathcal{O}_\ell := \left[C^\top, (CA)^\top, \dots, \left(CA^{\ell-1}\right)^\top\right]^\top$ has rank n_x. The set of all vectors $[u_{1:L}; y_{1:L}]$, where $[u_{1:L}; y_{1:L}]$ is an L-step I/O trajectory generated by the system $\mathcal{B}(A,B,C,D)$, is denoted by $\mathcal{B}_L(A,B,C,D)$.

Definition 1: A Hankel matrix of depth L constructed by

The first author received support from the Swiss National Science Foundation (SNSF) under the NCCR Automation project, grant agreement 51NF40_180545. The second author received support from the SNSF under the RISK project (Risk Aware Data-Driven Demand Response), grant number 200021 175627.

*: corresponding author

JS, YL and CNJ are with Automatic Laboratory, Ecole Polytechnique Federale de Lausanne, 1015 Lausanne, Switzerland. {jicheng.shi, yingzhao.lian, colin.jones}@epfl.ch

a vector-valued signal sequence $s := \{s_i\}_{i=1}^T$, $s_i \in \mathbb{R}^{n_s}$ is

$$\mathfrak{H}_L(s) := \begin{bmatrix} s_1 & s_2 & \dots & s_{T-L+1} \\ s_2 & s_3 & \dots & s_{T-L+2} \\ \vdots & \vdots & & \vdots \\ s_L & s_{L+1} & \dots & s_T \end{bmatrix} .$$

Given a sequence of historical input-output measurements $\{u_{d,i}, y_{d,i}\}_{i=1}^T$, the input sequence is called *persistently exciting* of order L if $\mathfrak{H}_L(u_d)$ is full row rank. By building the following stacked Hankel matrix $\mathfrak{H}_L(u_d, y_d) := \begin{bmatrix} \mathfrak{H}_L(u_d)^\top & \mathfrak{H}_L(y_d)^\top \end{bmatrix}^\top$, we state **Willems' Fundamental Lemma** as

Lemma 1: [11, Theorem 1] Consider a controllable linear system $\mathcal{B}(A, B, C, D)$ and assume $\{u_{d,i}\}_{i=1}^T$ is persistently exciting of order $L + n(\mathcal{B}(A, B, C, D))$. The condition $\text{colspan}(\mathfrak{H}_L(u_d, y_d)) = \mathcal{B}_L(A, B, C, D)$ holds.

In the rest of the paper, the subscript $_d$ marks a data point from the historical dataset collected offline, and L is reserved for the length of the system response.

The characterization of system response by Lemma 1 is used to develop a data-driven output prediction [12], [13]. In [12], the N_{pred}-step output prediction $\bar{y}_{t+1:t+N_{pred}}$ driven by an N_{pred}-step predicted input $u_{t+1:t+N_{pred}}$ is given by the solution to the following equations at time t:

$$\begin{bmatrix} \mathfrak{H}_{L,init}(u_d) \\ \mathfrak{H}_{L,init}(y_d) \\ \mathfrak{H}_{L,pred}(u_d) \end{bmatrix} g = \begin{bmatrix} u_{t-N_{init}+1:t} \\ y_{t-N_{init}+1:t} \\ u_{t+1:t+N_{pred}} \end{bmatrix} \tag{2a}$$

$$\mathfrak{H}_{L,pred}(y_d) g =: \bar{y}_{t+1:t+N_{pred}} . \tag{2b}$$

where $N_{init} + N_{pred} = L$ and $g \in \mathbb{R}^{T-L+1}$ is the solution to (2a). Two output sub-Hankel matrices are defined by

$$\mathfrak{H}_L(y_d) = \begin{bmatrix} \mathfrak{H}_{L,init}(y_d) \\ \mathfrak{H}_{L,pred}(y_d) \end{bmatrix} , \tag{3}$$

and each of them is of depth N_{init} and N_{pred} respectively. Similarly, the Hankel matrices $\mathfrak{H}_{L,init}(u_\mathbf{d})$ and $\mathfrak{H}_{L,pred}(u_\mathbf{d})$ are constructed. Last but not least, the estimation given by (2b) is unique if $N_{init} \geq l(\mathcal{B}(A, B, C, D))$. Specifically, this condition implies that $\{u_{t-N_{init}+1:t}, y_{t-N_{init}+1:t}\}$, the N_{init}-step input output sequences preceding the current point of time, can uniquely determine the underlying state x_t. Readers are referred to [12] for more details.

A. Problem Statement and Inspiration

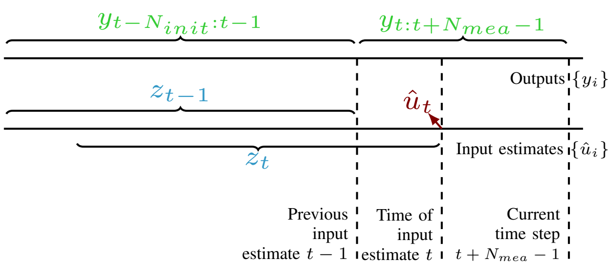


Fig. 1: Diagram of input estimation at time t. Input estimate $\hat{u}_t$ can be estimated after $y_{t+N_{mea}-1}$ is measured at time $t+N_{mea}-1$. Differently, output prediction $\bar{y}_{t+1}$ is computed immediately by (2), if given u_{t+1}.

We first assume that an offline I/O dataset $\{u_d, y_d\}$ is available. During online operation, the inputs are not measurable and are thus **unknown**, and this work studies the recursive estimation of the unknown inputs from the measured outputs. The recursive estimate is generated by the following linear system, a process we dub the 'input reconstruction method from outputs (IRO)'

$$\begin{aligned} z_t &= A_{IRO} z_{t-1} + B_{IRO} d_{t-1} , \\ \hat{u}_t &= [\mathbf{0} \; \mathbf{I}_{n_u}] z_t , \end{aligned} \tag{4}$$

where $z_t := [\hat{u}_{t-N_{init}+1}^\top, \dots, \hat{u}_t^\top]^\top$ is a vectorized N_{init}-step unknown input estimate, and $d_t := y_{t-N_{init}+1:t+N_{mea}}$ is the sequence of output measurements. The time instances correspondence of the IRO is depicted in Figure 1, and the input estimate is delayed by $N_{mea} - 1$. We leave the discussion about N_{mea} to Section III.

We call the IRO **stable** if $\lim\limits_{t\to\infty} \hat{u}_t - u_t \to 0$ for any initial guess z_0. Note that since z_t is the sequence of N_{init}-step unknown input estimates, it is reasonable to design an observable canonical form based IRO, such that the recursive estimator only updates the last unknown input in z_t (i.e. $\hat{u}_t$), and we term an IRO of this form a recursive IRO (R-IRO). Otherwise, it is called a moving-horizon IRO (MH-IRO).

The goal of this work is to design the IRO components (i.e. A_{IRO} and B_{IRO}) directly from data $\{u_d, y_d\}$. Inspired by the data-driven output prediction (2), it is reasonable to formulate a similar data-driven input estimation scheme:

$$\begin{bmatrix} \mathfrak{H}_{L,init}(u_d) \\ \mathfrak{H}_{L,init}(y_d) \\ \mathfrak{H}_{L,mea}(y_d) \end{bmatrix} g = \begin{bmatrix} u_{t-N_{init}:t-1} \\ y_{t-N_{init}:t-1} \\ y_{t:t+N_{mea}-1} \end{bmatrix} \tag{5a}$$

$$\mathfrak{H}_{L,est}(u_d) g =: \bar{u}_t , \tag{5b}$$

where $g \in \mathbb{R}^{T-L+1}$ is a solution to (5a). Additionally, sub-Hankel matrices $\mathfrak{H}_{L,init}(u_d)$, $\mathfrak{H}_{L,init}(y_d)$, $\mathfrak{H}_{L,mea}(y_d)$, $\mathfrak{H}_{L,mea}(u_d)$ follow a similar splitting definition to that in (3) with $N_{init} + N_{mea} = L$, and $\mathfrak{H}_{L,est}(u_d)$ denotes the first n_u rows of $\mathfrak{H}_{L,mea}(u_d)$. However, this scheme (5) is not implementable, as input measurements $u_{t-N_{init}:t-1}$ in (5a) are not available. The key idea of this work is to fit this scheme (5) into the general IRO structure (4).

Remark 1: In the rest of the paper, $\bar{u}_t$ indicates the input estimate by (5) given the actual $u_{t-N_{init}:t-1}$. $\hat{u}_t$ denotes the input estimate by (4) and z_t. N_{init} and L are user-defined, which can be different in (2) and (5). For the sake of consistency, we consider the same N_{init} and $N_{pred} = N_{mea}$ in the rest of this paper.

III. Data-Driven R-IRO

In this section, we present the proposed R-IRO and its design approach. We will first summarize its standard formulation and its stability property in Theorem 1, whose proof is later accomplished by Lemma 3 and Lyapunov conditions. The design of a standard R-IRO in Theorem 1 suffers from NP-hardness, and so we offer a tractable reformulation by the LMI tightening.

The key idea of the R-IRO formulation is to substitute $u_{t-N_{init}:t-1}$ in (5a) by its recursive input estimate $\hat{u}_{t-N_{init}:t-1} =: z_{t-1}$ in the IRO (4). For the sake of clarity, the notation in (5) is simplified using the notation

$$H := \begin{bmatrix} \mathfrak{H}_{L,init}(u_d) \\ \mathfrak{H}_{L,init}(y_d) \\ \mathfrak{H}_{L,mea}(u_d) \end{bmatrix} , \quad b := \begin{bmatrix} u_{t-N_{init}:t-1} \\ y_{t-N_{init}:t-1} \\ y_{t:t+N_{mea}-1} \end{bmatrix} , \quad H_u := \mathfrak{H}_{L,est}(y_d)$$

We state the set of data-driven R-IRO candidates by

\cnj{Some words here on why these are the right candidates would be helpful. A casual reader isn't going to be able to jump from the text to this set without some real thought. Or rather, just say clearly here that the development of this set is coming up in the rest of the section. As written, it sounds like the reader is supposed to see immediately that this is the right set.}

I think you've got your quantifiers backwards. As written, this describes a single A,B which must satisfy these conditions for all G. However, I think you want exists instead of forall, no?

$$\mathcal{U}_R := \left\{ \begin{matrix} A_{IRO}, \\ B_{IRO} \end{matrix} \middle| \begin{matrix} A_{IRO} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{(N_{init}-1)n_u} \\ & H_u G_u \end{bmatrix}, \\ B_{IRO} = \begin{bmatrix} \mathbf{0} \\ H_u G_y \end{bmatrix}, \ \forall\, [G_u, G_y] = G \in H^g \end{matrix} \right\} \quad (6)$$

where G_u and G_y partitions any generalized inverse G, and respectively consists of $N_{init} n_u$ and $(N_{init} + N_{mea}) n_y$ columns.

Assumption 1: The historical input signals $\{u_{d,i}\}_{i=1}^{T}$ are persistently exciting of order $N_{init} + N_{mea} + n(\mathcal{B}(A,B,C,D))$.

Theorem 1 (Stability of R-IRO): Let condition

$$\mathrm{Null}(H) \subseteq \mathrm{Null}(H_u)\ , \quad (7)$$

and Assumption 1 hold. The set of stable IRO in $\mathcal{U}_R$ is given by:

$$\left\{ A_{IRO}\ , B_{IRO} \in \mathcal{U}_R \middle| \exists W \succ 0, \begin{bmatrix} W & A_{IRO} W \\ W A_{IRO}^\top & W \end{bmatrix} \succ 0 \right\} \quad (8)$$

Do you mean the set of stable IRO in UR, or the set of stable IRO? i.e., is it possible that there exist stable IRO that are not in UR?

In the following, we will show how (6) and Thereom 1 are developed. Above all, Assumption 1 ensures that the Hankel matrix H constructed by $\{u_d, y_d\}$ is sufficiently informative such that Lemma 1 guarantees that $b \in \mathrm{colspan}(H)$. Therefore, the solution set to $Hg = b$ is non-empty, and it can be characterized as

$$\mathcal{T}(b) := \{g | g = Gb + \nu,\ G \in H^g,\ \nu \in \mathrm{Null}(H)\}\ . \quad (9)$$

Accordingly, $\bar{u}_t$ in (5b) is given by

$$\bar{u}_t = \{H_u g | g \in \mathcal{T}(b)\}\ . \quad (10)$$

However, the solution set (9) is not a singleton and therefore $\bar{u}_t$ is not necessarily unique. To ensure uniqueness, we give the following lemma.

Lemma 2: If Assumption 1 holds, the set (10) is a singleton if and only if the condition (7) holds.

Proof: ($\Rightarrow$) For any solutions $g_1,\ g_2 \in \mathcal{T}(b)$ to $Hg = b$, we have $Hg_1 - Hg_2 = H(g_1 - g_2) = 0$, which indicates $(g_1 - g_2) \in \mathrm{Null}(H)$ Therefore, by $\mathrm{Null}(H) \subseteq \mathrm{Null}(H_u)$, $H_u g_1 - H_u g_2 = 0$. Due to the arbitrariness of g_1 and g_2, $\bar{u}_t$ defined in (10) is a singleton. ($\Leftarrow$) For any $G \in H^g$, $\nu \in \mathrm{Null}(H)$, $H_u(Gb + \nu) - H_u Gb = 0$ because $\bar{u}_t$ by (10) is a singleton. This indicates $H_u \nu = 0, \forall \nu \in \mathrm{Null}(H)$ and therefore $\mathrm{Null}(H) \subseteq \mathrm{Null}(H_u)$. ∎

If the condition (7) holds, the effect of null space $\mathrm{Null}(H)$ in $\mathcal{T}(b)$ can be neglected. Next, by substituting $u_{t-N_{init}:t-1}$ in (10) by z_{t-1}, we formulate the input reconstruction by:

$$\forall\ G \in H^g,\ \hat{u}_t := H_u G [z_{t-1}^\top\ d_{t-1}^\top]^\top . \quad (11)$$

We can now see that the observable canonical form of this system results in the set of data-driven R-IRO candidates (6). In general, under the uniqueness condition (7), the solution to problem (5) inspires the R-IRO candidates in $\mathcal{U}_R$.

In the rest of this subsection, we will show the proof of Theorem 1. Note that an R-IRO defined by any element in set $\mathcal{U}_R$ (6) is not necessarily stable. To find a stable R-IRO, we first characterize its stability by the following lemma.

Lemma 3: Let Assumption 1 and condition (7) hold, an IRO in $\mathcal{U}_R$ is stable if and only if A_{IRO} is Schur.

Proof: If Assumption 1 and condition (7) holds, Lemma 1 and 2 guarantee that $\bar{u}_t = u_t$ in (10), with u_t being the real inputs. Therefore $\forall\ A_{IRO},\ B_{IRO} \in \mathcal{U}_R$, (10) is equivalent to

$$u_{t-N_{init}+1:t} = A_{IRO} u_{t-N_{init}:t-1} + B_{IRO} d_{t-1}\ ,$$

$$u_t = [\mathbf{0}\ \mathbf{I}_{n_u}] u_{t-N_{init}+1:t}\ .$$

Thus, we have

$$\lim_{t\to\infty} \hat{u}_t - u_t = \lim_{t\to\infty} [\mathbf{0}\ \mathbf{I}_{n_u}] A_{IRO}(z_{t-1} - u_{t-N_{init}:t-1})$$

$$= \lim_{t\to\infty} [\mathbf{0}\ \mathbf{I}_{n_u}] A_{IRO}^t (z_0 - u_{-N_{init}+1:0}).$$

The above equation converges to 0 if and only if A_{IRO} is Schur stable, and we conclude the proof. ∎

The Schur stability criterion can be validated via the following semidefinite program [20, Chapter 3.3]

$$\underset{\textsc{Schur stable}}{A_{IRO}} \Longleftrightarrow \left\{ \begin{matrix} \exists\, W \succ 0 \\ \begin{bmatrix} W & A_{IRO} W \\ W {A_{IRO}}^T & W \end{bmatrix} \succ 0. \end{matrix} \right. \quad (12)$$

This naturally leads to solution set (8) for the stable IRO, and we complete the proof of Theorem 1. In summary, the solution to (5) gives the candid a R-IRO structure in $\mathcal{U}_R$, and Lemma 3 helps us to find a stable element within $\mathcal{U}_R$, which is characterized by (8) in Theorem 1

Remark 2: The choices of N_{mea} for output measurement depends on the properties of matrices $\{B, C, D\}$ in the LTI dynamics, which intuitively reflects how soon all the entries of inputs can affect the output. For example, if D is full column rank, the effect from the input to output is instantaneous and thus N_{mea} can be set to one. For model-based methods, a discussion about N_{mea} can be found in [7] and [21]. The condition (7) in Lemma 2 gives a data-driven criterion of N_{mea} selection, which intuitively states that the variation in input will always change the output, as any $g \notin \mathrm{Null}(H_u)$ is not in $\mathrm{Null}(H)$.

A. Design of data-driven R-IRO

Theorem 1 gives a design procedure for a data-driven R-IRO via the search of a feasible point in (8). However, this optimization problem is NP-hard due to the bilinear matrix (BMI) inequality in (8) [22]. In the rest of this section, we will tighten this BMI into a tractable the matrix inequality (LMI) [23] the characterize the set of the generalized inverse H^g in U_R via singular value decomposition (SVD).

Would the algebra become simpler if we just partitioned the U and V matrices? i.e., suppose that U = [U1 U2; U2 U3], then Hg would be
V1*inv(S)*U3' + V3*F*U3'
?
Would this also simplify / clarify the N1, N2 and N3 computations?

1) Characterization of Generalized Inverse: Denote the SVD of matrix H by $H = U\begin{bmatrix} S & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}V^\top$, with $S \in \mathbb{R}^{n_S \times n_S}$ containing all the positive singular values. Then the generalized inverse is characterized by

$$H^g = \left\{ G \;\middle|\; \begin{array}{l} V\left(\begin{bmatrix} S^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} + F\right)U^\top \\ F \in \mathbb{R}^{m_H \times n_H},\ [I_{n_S}\ \mathbf{0}]F\begin{bmatrix} I_{n_S} \\ \mathbf{0}\end{bmatrix} = \mathbf{0} \end{array}\right\}, \quad (13)$$

where F is a any matrix of shape H whose upper-left block of size $n_S \times n_S$ is zero. For the sake of clarity, we characterize an element in H^g by $G(F)$ such that

$$G(F) := V\left(\begin{bmatrix} S^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} + F\right)U^\top .$$

The set (13) is indeed the set of generalized inverse as

$$\begin{aligned} HG(F)H &= U\begin{bmatrix} S & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\left(\begin{bmatrix} S^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} + F\right)\begin{bmatrix} S & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}V^\top \\ &= U\begin{bmatrix} S & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}V^\top = H \end{aligned}$$

2) LMI Tightening: Before going into details, we would first intuitively explain the idea behind the design procedure. Recall the idea behind Lemma 2, we can see that the design of A_{IRO} lies in the selection of the null space of the matrix H such that the set (10) is still unique, and the matrix A_{IRO} is Schur stable. Hence, we only need to focus on the null space of H, which motivates the following LMI reformulation. Based on the characterization of H^g in (13), any A_{IRO} in our feasible set $\mathcal{U}_R$ is accordingly parametrized by matrix F such that

$$\begin{aligned} A_{IRO}(F) &= \begin{bmatrix} \mathbf{0} & \mathbf{I}_{(N_{init}-1)n_u} \\ \multicolumn{2}{c}{H_u V(\begin{bmatrix} S^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} + F)U^\top \begin{bmatrix} \mathbf{I}_{N_{init}n_u} \\ \mathbf{0}\end{bmatrix}} \end{bmatrix} \\ &= N_1 + N_2 F N_3, \\ N_2 &:= \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{n_u}\end{bmatrix} H_u V, \quad N_3 = U^\top \begin{bmatrix} \mathbf{I}_{N_{init}n_u} \\ \mathbf{0}\end{bmatrix} \\ N_1 &:= \begin{bmatrix} \mathbf{0} & \mathbf{I}_{(N_{init}-1)n_u} \\ \multicolumn{2}{c}{\mathbf{0}} \end{bmatrix} + N_2 \begin{bmatrix} S^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} N_3 , \end{aligned} \quad (14)$$

To enable the LMI reformulation, we define

$$T_1 = [\mathbf{0}\ \ \mathbf{I}_{n_H - n_S}] \in \mathbb{R}^{(n_H - n_S)\times n_H} , \quad (15)$$

and we denote $r = \mathrm{rank}(T_1 N_3)$. Regarding the definition of generalized inverse and (14), the operation T_1N_3 selects the components in U related to $\mathrm{Null}(H)$. Followed by this, we define $T_2 = [\mathbf{I}_r\ \mathbf{0}]E \in \mathbb{R}^{r\times(n_H - n_S)}$, where E is the multiplication of elementary operations to execute Gauss-Jordan Elimination for T_1N_3. In summary, the operation $T_2T_1N_3$ generates the subspace of U related to the $\mathrm{Null}(H)$, and based on the aforementioned discussion, the design of A_{IRO} lies within this space, which leads to the following LMI tightening.

Lemma 4: The BMI constraint in (8) is satisfied if $\exists\, N \in \mathbb{R}^{m_H \times r}, M \in \mathbb{R}^{r\times r}$ and $W \in \mathbb{R}^{N_{init}nu \times N_{init}nu} \succeq 0$ such that $F = NM^{-1}T_2T_1$ and

$$\left\{ \begin{array}{ll} \begin{bmatrix} W & \begin{array}{c} N_1W + \\ N_2NT_2T_1N_3 \end{array} \\ \begin{array}{c} WN_1{}^T + \\ (N_2NT_2T_1N_3)^T \end{array} & W \end{bmatrix} \succ 0 & (16a) \\ T_2T_1N_3W = MT_2T_1N_3 & (16b) \end{array}\right.$$

For any feasible solution F, the corresponding A_{IRO}, $B_{IRO} \in \mathcal{U}_R$ are later reconstructed by setting $G \in H^g$ to $G(F)$.

Proof: The first n_S columns of F are zeros, because $F = NM^{-1}T_2T_1$ and $T_1 = [\mathbf{0}\ \ \mathbf{I}_{n_H - n_S}]$. Hence, F satisfies (13) and gives a generalized inverse $G(F)$.

The rest of the proof is similar to [23, Theorem 1]. By definition of T_2, matrix $T_2T_1N_3$ is full row rank. The left-hand-sied of condition (16b) is therefore full rank as $W \succ 0$, which further ensures that M is also full rank. Therefore, M^{-1} exists and we get $T_2T_1N_3 = M^{-1}T_2T_1N_3W$ from (16b). Then we get the BMI in (8) from (16a) by

$$\begin{aligned} N_1W + N_2NT_2T_1N_3 &= N_1W + N_2NM^{-1}T_2T_1N_3W \\ &\overset{(a)}{=} N_1W + N_2FN_3W = A_{IRO}(F)W , \end{aligned}$$

Comment on how / why this is a conservative tightening.

IV. Data-Driven MH-IRO

Recall that an R-IRO only updates the most recent unknown input in z_t, i.e. $\hat{u}_t$. Similar to the concept used for a Luenberger observer [24], the key idea behind a data-driven MH-IRO is to enable the correction update of the $\hat{u}_{t-N_{init}+1:t}$ estimate, i.e. z_t, by the error between the actual measurement of y_t and its data-driven predictive estimate $\hat{y}_t$.

Recall the data-driven prediction problem (2) in Section II. We focus on the prediction of y_t, replace $u_{t-N_{init}+1:t}$ by z_t and define following matrices for the sake of clarity,

$$\tilde{H} := \begin{bmatrix} \mathfrak{H}_{L,init}(u_d) \\ \mathfrak{H}_{L,est}(u_d) \\ \mathfrak{H}_{L,init}(y_d) \end{bmatrix}, \ H_y := \begin{bmatrix}\mathbf{I}_{n_y} & \mathbf{0}\end{bmatrix}\mathfrak{H}_{L,mea}(y_d)$$

$$\begin{aligned} \tilde{b} &:= \begin{bmatrix} \hat{u}_{t-N_{init}:t-1} \\ \hat{u}_t \\ y_{t-N_{init}:t-1} \end{bmatrix} \overset{(a)}{=} \begin{bmatrix} z_{t-1} \\ H_u(G_u z_{t-1} + G_y d_{t-1}) \\ [I_{n_yN_{init}}\ \mathbf{0}]d_{t-1} \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} I & \mathbf{0} \\ H_uG_u & H_uG_y \\ \mathbf{0} & [I_{n_yN_{init}}\ \mathbf{0}] \end{bmatrix}}_{P(G)} \begin{bmatrix} z_{t-1} \\ d_{t-1} \end{bmatrix}, \ \forall\, [G_u\ G_y] = G \in H^g \end{aligned}$$

where (a) follows (11) and $P(G)$ is introduced for the sake of compactness. Then, similar to (11), the corresponding output prediction $\hat{y}_t$ is defined by

$$\begin{aligned} \forall\, \tilde{G} \in \tilde{H}^g,\ \hat{y}_t :=& H_y\tilde{G}\tilde{b} \\ =& H_y\tilde{G}P(G)\begin{bmatrix} z_{t-1} \\ d_{t-1} \end{bmatrix} \end{aligned} \quad (18)$$

Under Assumption 1 and $N_{init} \geq l(\mathcal{B}(A,B,C,D))$, the Fundamental Lemma 1 and Lemma 2 guarantees this equality holds for the actual output y_t with respect to the actual but unknown previous input sequence $u_{t-N_{init}:t-1}$

$$\forall\, \tilde{G} \in \tilde{H}^g,\ y_t = H_y\tilde{G}P(G)\begin{bmatrix} u_{t-N_{init}:t-1} \\ d_{t-1} \end{bmatrix} \quad (19)$$

Following a Luenberger observer style design, the observer will have the following structure with $\tilde{A}_{IRO}$, $\tilde{B}_{IRO} \in \mathcal{U}_R$:

$$z_t = \tilde{A}_{IRO} z_{t-1} + \tilde{B}_{IRO} d_{t-1} + L(y_t - \hat{y}_t) \ ,$$

where $L \in \mathbb{R}^{n_u N_{init} \times n_y}$ is a design parameter, $\hat{y}_t$ is given in (18) and y_{t+1} is always an entry of d_t as $N_{mea} \geq 1$ with

$$y_t = \overbrace{[\underbrace{\mathbf{0}}_{(a)} \ \mathbf{I}_{n_y} \ \underbrace{\mathbf{0}}_{(b)}]}^{T_y} d_{t-1} \ .$$

The term (a) is of $N_{init} n_y$ columns and the term (b) is of $(N_{mea}-1)n_y$ columns, and this linear mapping is denoted by T_y. Hence, $\forall \ \tilde{A}_{IRO}, \ \tilde{B}_{IRO} \in \mathcal{U}_R, \ \tilde{G} \in \tilde{H}^g, \ G \in H^g$, the components of a data-driven MH-IRO can be written as:

$$A_{IRO} = \tilde{A}_{IRO} - LH_y\tilde{G}P(G)\begin{bmatrix}\mathbf{I}_{n_y N_{init}} \\ \mathbf{0}\end{bmatrix} \tag{20a}$$

$$B_{IRO} = \tilde{B}_{IRO} + LT_y - LH_y\tilde{G}P(G)\begin{bmatrix}\mathbf{0} \\ \mathbf{I}_{n_y(N_{init}+N_{mea})}\end{bmatrix} \tag{20b}$$

The following Theorem summarizes the stability of a data-driven MH-IRO.

Theorem 2 (Stability of MH-IRO): Let Assumption 1 and condition (7) hold. For any $\tilde{A}_{IRO}$,$\tilde{B}_{IRO} \in \mathcal{U}_R$, $\tilde{G} \in \tilde{H}^g$, $G \in H^g$, the data-driven MH-IRO in (20) is stable if $\tilde{A}_{IRO} - LH_y\tilde{G}P(G)\begin{bmatrix}\mathbf{I}_{n_y N_{init}} \\ \mathbf{0}\end{bmatrix}$ is Schur stable.

Proof: Similar to the proof of Lemma 3. ■

In conclusion, the design process and operation of R-IRO and MH-IRO are summarized in Algorithm 1.

Algorithm 1 Design and operation a data-driven IRO

Given historical signals $\{u_{d,i}, y_{d,i}\}_{i=1}^{T}$.

1) Choose a large N_{init}, Choose Nmea > 1 such that condition (7) holds. until the condition (7) h
2) Build the IRO in the form (4) by either:
   a) R-IRO: Compute G by either (8) or (16). Compute the components in (6).
   b) MH-IRO: Choose any $G \in H^g$, $\tilde{G} \in \tilde{H}^g$. Design L such that the stability condition in Theorem 2 holds. Compute the components in (20).
3) From $t = 0$, choose arbitrary z_0 and repeatedly compute (4) to output $\hat{u}_t$.

Remark: … something about how selecting the smallest Nmea s.t. (7) holds is a good idea.

Remark 3: The design methods by Lyapunov condition (8), LMI formulation (16) and MH-IRO in Theorem (20) do not guarantee the existence of a data-driven IRO for any system. The existence problem of an IRO has been explored in a model-based setup, which shows that the existence is related to the system dynamic $\mathcal{B}(A, B, C, D)$ [1], [6]. However, the existence problem within a data-driven setup is still unclear and remains future work.

Remark 4: In comparison with the data-driven R-IRO, we observed that the data-driven MH-IRO is more robust to measurement noise contaminated data, because it does not require any construction of the null space, which may be sensitive to measurement noise [25].

V. Simulation and Experimental Validation

A. Simulation

We consider the following **unstable** discrete-time LTI dynamics:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{c|c} \begin{bmatrix} 0.9 & 1.4 & 0.2 \\ 0.5 & 1.5 & 1.5 \\ 1.6 & 0.6 & 0.4 \end{bmatrix} & \begin{bmatrix} 0.5 & 1.0 \\ 0.9 & 0.3 \\ 0.4 & 0.3 \end{bmatrix} \\ \hline \begin{bmatrix} 1.5 & 1.0 & 1.4 \\ 0.6 & 0.3 & 0.3 \end{bmatrix} & \gamma \begin{bmatrix} 1.3 & 1.8 \\ 0.4 & 0.7 \end{bmatrix} \end{array}\right]$$

Where did this come from? Is it a benchmark used in another paper?

The proposed schemes are compared with two indirect UIO-based methods, including a model-based UIO [6] and a data-driven UIO (DD-UIO) [19]. The historical I/O data are generated by a 50-step trajectory excited by random inputs. For both IROs, we set $N_{init} = 5$ and the initial guess to $z_0 = [0 \ 0 \ \ldots 0 \ 0]^\top$. The same dataset is used to identify a state-space model. In particular, this model is directly used to design the model-based UIO, and it is used to recursively update the state estimate used in DD-UIO by a Luenberger observer. Recall remark 2, different γ results in different N_{mea}. By checking condition (7), we consider $\gamma = 1$ with $N_{mea} = 1$ and $\gamma = 0$ with $N_{mea} = 2$. The same choices of N_{mea} are used in the benchmark UIOs.

If there's no noise, and you have a sufficiently excited system, isn't this the same as just using the true system matrices?

The results of $\hat{u}_t(i)$ by R-IRO are plotted in Figure 2(a). The estimation error $du_t(i) = u_t(i) - \hat{u}_t(i)$ is given in Figure 2(b) and (c), where all the design schemes show fast convergence in the estimate even though the underlying dynamic are unstable.

Remark 5: Unknown input observers (UIOs) are also widel[illegible]known input reconstruction. Without the direct system states measurement [illegible]t, the design of a UIO depends on a state-space model, whose identification requires singular value decomposition or QR decomposition [26]. Hence, the computational complexity of designing a UIO is at least as high as the proposed schemes. Furthermore, both UIO and the proposed schemes have the same linear asymptotic convergence speed as linear estimators. In general, the proposed methods are preferable to the UIO-based methods, because the engineering effort of system identification is saved.

This could be a contentious statement… ;)

B. Experiment

This experiment is carried out on a whole building, named the Polydome, on the EPFL campus, and we estimate the number of occupants by an indoor CO2 level measurement. Although the building dynamics are nonlinear due to the ventilation system, it has a good linear approximation when the ventilation flow rate is constant [27]. Under the assumption that the CO2 generation rate per person doing office work is relatively constant, the proposed schemes in this work are feasible. The offline dataset contains indoor temperature, weather condition, heat pump power, CO2 level, and occupant number recorded by manual headcount (i.e., online measurement is not affordable). The indoor CO2 level is measured as the averaged value from four air quality sensors, whose installation locations are shown in Figure 3. Data from five weekdays are used to build the Hankel matrix,

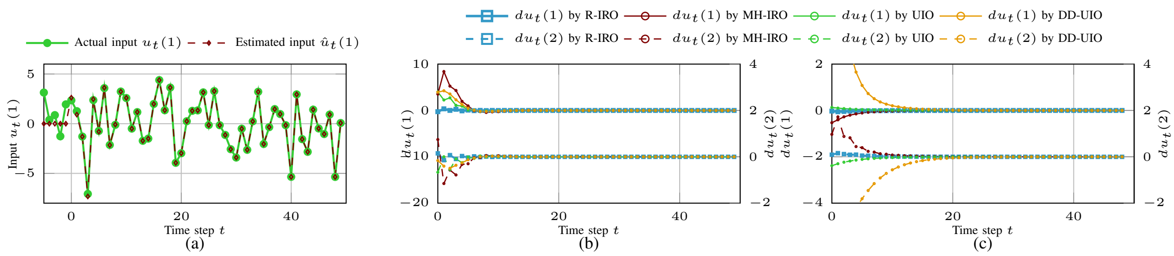


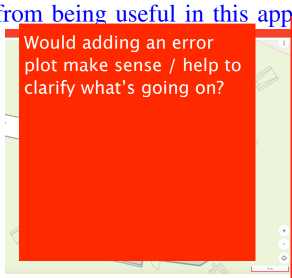
Fig. 2: Simulation: input estimation by R-IRO, MH-IRO, UIO and DD-UIO. (a): R-IRO, $\gamma = 1$, input $u_t(1)$. Two IROs, UIO and DD-UIO: (b): $\gamma = 1$. (c): $\gamma = 0$.

and the proposed data-driven MH-IRO[1] is compared with linear regression (LR), Gaussian process regression (GR), model-based UIO and DD-UIO by another five-weekday data. Note outside the office hours, i.e., between $7:00$ PM and $7:00$ AM, we enforce $\hat{u}_t = 0$ within this time interval to improve the estimate. The results are plotted in Figure 4. From the top plot, one can see that the proposed MH-IRO scheme is better than LR and slightly worse than GR in terms of mean absolute error (MAE). However, the UIR better tracks the occupancy trajectory while GR shows significant fluctuations in its estimates. Meanwhile, the performance of the UIO and the DD-UIO is even lower than the linear regression's (bottom plot Figure 4) in terms of MAE. Additionally, their strong fluctuation in the estimates prevents them from being useful in this application.

How come we don't have plots for R-IRO and MH-IRO seperately here?

Would adding an error plot make sense / help to clarify what's going on?

A brief statement about the design of the GPR and LR would make sense, since there's a lot of design decisions and tuning parameters there. Also a reference to it, since I don't think it's been mentioned previously?



Use the same scales on the two plots if they're showing the same thing.

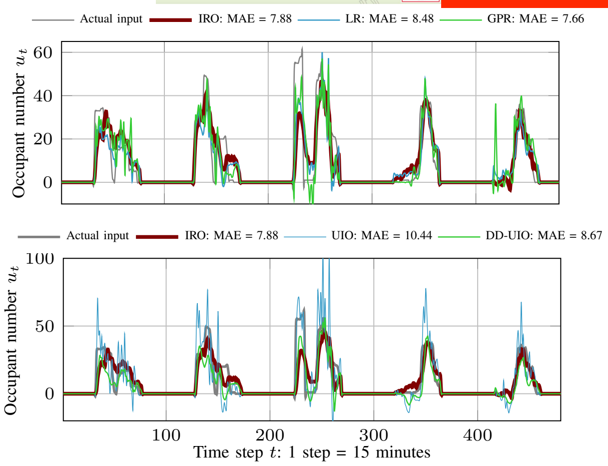


Fig. 4: Comparison of occupant number estimation. Top: MH-IRO, LR and GPR. Bottom: MH-IRO, UIO and DD-UIO. Mean absolute error (MAE) is computed for the data during office hours.

VI. Conclusions

This work proposes two data-driven IRO design schemes based on a Lyapunov condition and a Luenberger-observer-type feedback. The stability of the proposed schemes are discussed, and their efficacy is validated by numerical simulations and a real-world experiment of occupancy estimation.

[1]The R-IRO does not give good performance in this experiment due to the measurement noise within the data and the nonlinearity of the dynamics.

References

[1] M. Hou and R. J. Patton, "Input observability and input reconstruction," *Automatica*, vol. 34, no. 6, pp. 789–794, 1998.

[2] A. Ansari, "Input and state estimation for discrete-time linear systems with application to target tracking and fault detection," Ph.D. dissertation, 2018.

[3] R. Rajamani, Y. Wang, G. D. Nelson, R. Madson, and A. Zemouche, "Observers with dual spatially separated sensors for enhanced estimation: Industrial, automotive, and biomedical applications," *IEEE Control Systems Magazine*, vol. 37, no. 3, pp. 42–58, 2017.

[4] Z. Chen, C. Jiang, and L. Xie, "Building occupancy estimation and detection: A review," *Energy and Buildings*, vol. 169, pp. 260–270, 2018.

[5] F. Zhu, "State estimation and unknown input reconstruction via both reduced-order and high-order sliding mode observers," *Journal of Process Control*, vol. 22, no. 1, pp. 296–302, 2012.

[6] M. E. Valcher, "State observers for discrete-time linear systems with unknown inputs," *IEEE Trans. Autom. Control*, vol. 44, no. 2, pp. 397–401, 1999.

[7] S. Sundaram and C. N. Hadjicostis, "Delayed observers for linear systems with unknown inputs," *IEEE Trans. Autom. Control*, vol. 52, no. 2, pp. 334–339, 2007.

[8] S. Gillijns and B. De Moor, "Unbiased minimum-variance input and state estimation for linear discrete-time systems," *Automatica*, vol. 43, no. 1, pp. 111–116, 2007.

[9] A. Ansari and D. S. Bernstein, "Deadbeat unknown-input state estimation and input reconstruction for linear discrete-time systems," *Automatica*, vol. 103, pp. 11–19, 2019.

[10] A. Kheder, K. Benothman, M. Benrejeb, and D. Maquin, "State and unknown input estimation via a proportional integral observer with unknown inputs," in *9th International conference on Sciences and Techniques of Automatic control & computer engineering, STA'2008*, 2008, p. CDROM.

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

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

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

[14] 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.

[15] I. Markovsky and P. Rapisarda, "On the linear quadratic data-driven control," in *2007 Eur. Control Conf. (ECC)*. IEEE, 2007, pp. 5313–5318.

[16] J. Berberich, J. Köhler, M. A. Müller, and F. Allgöwer, "Data-driven model predictive control with stability and robustness guarantees," *IEEE Trans. Autom. Control*, vol. 66, no. 4, pp. 1702–1717, 2020.

[17] Y. Lian, J. Shi, M. P. Koch, and C. N. Jones, "Adaptive robust data-driven building control via bi-level reformulation: an experimental result," *arXiv preprint arXiv:2106.05740*, 2021.

[18] Y. Lian and C. N. Jones, "Nonlinear data-enabled prediction and control," in *Learning for Dynamics and Control*. PMLR, 2021, pp. 523–534.

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

[20] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.

[21] J. Jin, M.-J. Tahk, and C. Park, "Time-delayed state and unknown input observation," *International Journal of Control*, vol. 66, no. 5, pp. 733–746, 1997.

[22] O. Toker and H. Ozbay, "On the np-hardness of solving bilinear matrix inequalities and simultaneous stabilization with static output feedback," in *Proceedings of 1995 Amer. Control Conf. -ACC'95*, vol. 4. IEEE, 1995, pp. 2525–2526.

[23] C. A. Crusius and A. Trofino, "Sufficient lmi conditions for output feedback control problems," *IEEE Trans. Autom. Control*, vol. 44, no. 5, pp. 1053–1057, 1999.

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

[25] R.-C. Li, "Relative perturbation theory: Ii. eigenspace and singular subspace variations," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 471–492, 1998.

[26] P. Van Overschee and B. De Moor, "A unifying theorem for three subspace system identification algorithms," *Automatica*, vol. 31, no. 12, pp. 1853–1864, 1995.

[27] D. Calì, P. Matthes, K. Huchtemann, R. Streblow, and D. Müller, "Co2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings," *Building and Environment*, vol. 86, pp. 39–49, 2015.