University of Groningen

Beyond Persistent Excitation

van Waarde, Henk

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Beyond Persistent Excitation: Online Experiment Design for Data-Driven Modeling and Control

Henk J. van Waarde[ORCID]

Abstract—**This letter presents a new experiment design method for data-driven modeling and control. The idea is to select inputs *online* (using past input/output data), leading to desirable rank properties of data Hankel matrices. In comparison to the classical persistency of excitation condition, this online approach requires less data samples and is even shown to be completely sample efficient.**

Index Terms—**Identification, linear systems.**

I. INTRODUCTION

RECENTLY, there has been an increasing interest in the direct design of controllers using data [1]–[12]. Several contributions study how controllers can be obtained from a given batch of (informative) data, even in the presence of noise [4], [6], [7] and for classes of nonlinear systems [11], [12]. The question of how to *obtain* such informative data sets, however, is largely open. For data-driven control to become an end-to-end solution, there is a need for new *experiment design* methods to empower the data-based design. This is true especially for settings including noise and nonlinear dynamics. However, even for linear systems with exact data, current experiment design methods are not sample efficient.

In this letter we will explore a new idea for designing experiments for data-driven modeling and control. Experiment design is a classical problem that has been mostly studied in the parametric identification literature. An established idea is to optimize a measure of the expected accuracy of the parameter estimates subject to input power constraints [13], [14]. This problem is usually tackled in the frequency domain and convex formulations have been provided in [15]. The dual problem of finding the "least costly" input achieving a fixed level of parameter accuracy has also been studied [16], in a closed-loop setting.

Less results are known in the context of non-parametric methods. In this area, a state-of-the-art result for linear time-invariant systems is Willems *et al.*'s fundamental lemma [17]. The idea behind this method is to select an input that is persistently exciting, which implies that a Hankel matrix of measured inputs and outputs satisfies a rank condition. This

rank property is important, since it guarantees that *all* trajectories of the system can be parameterized in terms of the measured trajectory. Essentially, the Hankel matrix of measured inputs and outputs serves as a non-parametric model of the system. This idea is simple yet powerful, and has been successfully employed in a number of recent publications on data-driven simulation and control, see, e.g., [1]–[3].

Despite its elegance and clear impact on data-driven methods, the fundamental lemma has some limitations.

1) It is not *sample efficient* in the sense that it uses more data than strictly necessary. This is because the lemma works with inputs that are persistently exciting of a certain order, which imposes a conservative lower bound on the number of samples. Large data sets are challenging from a computational point of view, especially for the real-time implementation of controllers, see, e.g., [18].

2) It is not applicable to *noisy data*. A possible approach to deal with this is to perform approximate data-driven simulation and control in a maximum likelihood framework [19], for which an experiment design method was developed in [20]. Another line of work establishes data-driven control design methods with guaranteed stability and performance in the presence of bounded process noise [4], [6], [7]. However, for this, experiment design methods are missing and it is unclear whether persistency of excitation is helpful in this context.

The purpose of this letter is to introduce another angle of attack to experiment design. In contrast to Willems' fundamental lemma, we do not use persistently exciting inputs, but instead design the inputs *online*. This means that at each time step, the input is selected on the basis of inputs and outputs that have been collected at previous time steps. Such an online approach is natural because the first samples of an experiment already contain valuable partial information about the system, which can be exploited in the design of the remaining samples. Online approaches appear in various contexts in systems and control. We mention iterative feedback tuning [21] that utilizes repeated (closed-loop) experiments, contributions to dual control [22]–[24] that trade-off exploration and exploitation, and adaptive experiment design [25], [26] where optimal experiment design is combined with adaptive parameter estimation. Nonetheless, in the context of Willems' lemma, online input design has not received any attention.

As our main contributions, we provide online experiment design methods for both input/state and input/output systems. In both settings, we formally prove that the methods are sample efficient. This completely resolves the problem in 1), and thus shows an advantage of online experiment design over

the classical persistency of excitation condition. Analogous to the fundamental lemma, our technical results are presented for noise-free data. The extension to noisy data is left for future work, and some ideas for this will be discussed in Section IV.

A. Notation and Terminology

The *left kernel* of a real matrix M is denoted by leftker M. Consider a signal $f : \mathbb{Z} \to \mathbb{R}^\bullet$ and let $i, j \in \mathbb{Z}$ be integers such that $i \leq j$. We denote by $f_{[i,j]}$ the restriction of f to the interval $[i, j]$, that is,

$$f_{[i,j]} := \begin{bmatrix} f(i)^\top & f(i+1)^\top & \cdots & f(j)^\top \end{bmatrix}^\top.$$

With slight abuse of notation, we will also use the notation $f_{[i,j]}$ to refer to the sequence $f(i), f(i+1), \ldots, f(j)$. Let k be a positive integer such that $k \leq j - i + 1$ and define the *Hankel matrix* of depth k, associated with $f_{[i,j]}$, as

$$\mathcal{H}_k(f_{[i,j]}) := \begin{bmatrix} f(i) & f(i+1) & \cdots & f(j-k+1) \\ f(i+1) & f(i+2) & \cdots & f(j-k+2) \\ \vdots & \vdots & & \vdots \\ f(i+k-1) & f(i+k) & \cdots & f(j) \end{bmatrix}.$$

Note that the subscript k refers to the number of block rows of $\mathcal{H}_k$. The sequence $f_{[i,j]}$ is called *persistently exciting of order k* if $\mathcal{H}_k(f_{[i,j]})$ has full row rank.

II. RANK CONDITIONS ON DATA MATRICES

Consider the linear time-invariant (LTI) system

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t) \tag{1a}$$

$$\mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{u}(t), \tag{1b}$$

where $\mathbf{x} \in \mathbb{R}^n$ denotes the state, $\mathbf{u} \in \mathbb{R}^m$ is the input and $\mathbf{y} \in \mathbb{R}^p$ is the output. Throughout this letter, we will assume that (1) is *minimal*, i.e., the pair (A, B) is controllable and (C, A) is observable. The *lag* ℓ of (1) can be defined as the smallest integer i for which the observability matrix

$$\mathcal{O}_i := \begin{bmatrix} C^\top & (CA)^\top & (CA^2)^\top & \cdots & (CA^{i-1})^\top \end{bmatrix}^\top \tag{2}$$

has rank n. Now, let $(u_{[0,T-1]}, y_{[0,T-1]})$ be an input/output trajectory of (1) of length $T \geq \ell$. In addition, let $L \geq \ell$ be an integer. The main goal of this letter is to provide a new method to design the input sequence $u_{[0,T-1]}$ so that the resulting input/output Hankel matrix

$$\begin{bmatrix} \mathcal{H}_L(y_{[0,T-1]}) \\ \mathcal{H}_L(u_{[0,T-1]}) \end{bmatrix} = \begin{bmatrix} y(0) & y(1) & \cdots & y(T-L) \\ \vdots & \vdots & & \vdots \\ y(L-1) & y(L) & \cdots & y(T-1) \\ u(0) & u(1) & \cdots & u(T-L) \\ \vdots & \vdots & & \vdots \\ u(L-1) & u(L) & \cdots & u(T-1) \end{bmatrix} \tag{3}$$

has rank $n+mL$. This rank condition plays a fundamental role in modeling and control using data. Indeed, it implies that any length-L input/output trajectory of (1) can be obtained from (3). More specifically, if (3) has rank $n + mL$ then $(\bar{u}_{[0,L-1]}, \bar{y}_{[0,L-1]})$ is an input/output trajectory of (1) if and only if

$$\begin{bmatrix} \bar{y}_{[0,L-1]} \\ \bar{u}_{[0,L-1]} \end{bmatrix} \in \text{im} \begin{bmatrix} \mathcal{H}_L(y_{[0,T-1]}) \\ \mathcal{H}_L(u_{[0,T-1]}) \end{bmatrix}.$$

This parameterization of input/output trajectories has been used extensively to simulate and control dynamical systems, see, e.g., [1]–[3]. If $L > \ell$, the rank condition on (3) even implies that *all* input/output trajectories of (1) (not just those of length L) can be obtained from $(u_{[0,T-1]}, y_{[0,T-1]})$, see [27]. In this case, the input/output behavior of (1) is *identifiable* [28] and (A, B, C, D) can be computed up to similarity transformation, e.g., using subspace methods [29].

A notable special case is that of *full state measurement*, i.e., $\mathbf{y}(t) = \mathbf{x}(t)$ and $L = \ell = 1$. In this case, (3) reduces to

$$\begin{bmatrix} \mathcal{H}_1(x_{[0,T-1]}) \\ \mathcal{H}_1(u_{[0,T-1]}) \end{bmatrix} = \begin{bmatrix} x(0) & x(1) & \cdots & x(T-1) \\ u(0) & u(1) & \cdots & u(T-1) \end{bmatrix}. \tag{4}$$

If (4) has rank $n+m$, the input/state trajectory $(u_{[0,T-1]}, x_{[0,T]})$ fully captures the behavior of (1a) which allows the unique identification of A and B.

A. Recap of Willems et al.'s Fundamental Lemma

Willems *et al.*'s fundamental lemma [17] is an important *experiment design* result. It reveals that the rank condition on (3) is satisfied if the input sequence is chosen to be persistently exciting. In a state-space setting, this result can be stated as follows [17], [30].

Proposition 1: Consider the minimal system (1). Let $(u_{[0,T-1]}, y_{[0,T-1]})$ be an input/output trajectory of (1) and let $L \geq \ell$. If the input $u_{[0,T-1]}$ is persistently exciting of order $n + L$ then (3) has rank $n + mL$.

We note that persistency of excitation of order $n+L$ requires at least $T \geq (m + 1)(n + L) - 1$ data samples.

III. ONLINE INPUT DESIGN

The main contribution of this letter will be to provide a new input design technique to guarantee the rank property of (3). In contrast to Willems' fundamental lemma, this method will not rely on applying inputs that are persistently exciting of order $n + L$. Rather, each input $u(t)$ is selected *online*, by making use of the previous samples $y(0), y(1), \ldots, y(t-1)$ and $u(0), u(1), \ldots, u(t-1)$, see Figure 1.

A. Input Design for Input/State Systems

To explain the idea, we will start with the simplest setting of *input/state* data collected from system (1a). In this setting, the purpose is to design a sequence of inputs $u(0), u(1), \ldots, u(T-1)$ so that the matrix (4) has full row rank $n+m$. As mentioned, we opt for the *online* design of the inputs. Specifically, this means that for the design of $u(t)$ we make use of the collected states[1] $x(0), x(1), \ldots, x(t)$ and inputs $u(0), u(1), \ldots, u(t-1)$. The following theorem shows how such inputs and a time horizon T can be designed.

Theorem 1: Consider the controllable system (1a). Define $T := n + m$. Select a nonzero $u(0) \in \mathbb{R}^m$, and design the input $u(t)$ (for $t = 1, 2, \ldots, T - 1$) as follows:

- If $x(t) \notin \text{im} \, \mathcal{H}_1(x_{[0,t-1]})$, select $u(t) \in \mathbb{R}^m$ arbitrarily.
- If $x(t) \in \text{im} \, \mathcal{H}_1(x_{[0,t-1]})$ there exists a $\xi \in \mathbb{R}^n$ and a nonzero $\eta \in \mathbb{R}^m$ such that

$$\begin{bmatrix} \xi^\top & \eta^\top \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-1]}) \\ \mathcal{H}_1(u_{[0,t-1]}) \end{bmatrix} = 0. \tag{5}$$

---

[1]In this setting, we use $x(t)$ in the design since it is independent of $u(t)$.
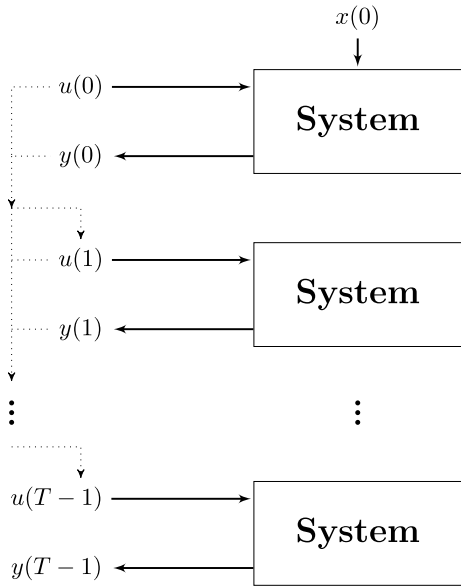
Fig. 1. Schematic representation of the online input design approach.

In this case, select $u(t)$ such that $\xi^\top x(t) + \eta^\top u(t) \neq 0$. Then, we have that

$$\text{rank} \begin{bmatrix} \mathcal{H}_1(x_{[0,T-1]}) \\ \mathcal{H}_1(u_{[0,T-1]}) \end{bmatrix} = n + m. \tag{6}$$

We will not provide a proof of Theorem 1 at this stage, since the statement will follow from the more general Theorem 2 that will be proven in Section III-B. Instead, we will explain the main ideas behind the result. The essence of Theorem 1 is that the proposed input sequence increases the rank of the input/state Hankel matrix *at every time step*. To be specific, the input sequence guarantees that

$$\text{rank} \begin{bmatrix} \mathcal{H}_1(x_{[0,t]}) \\ \mathcal{H}_1(u_{[0,t]}) \end{bmatrix} > \text{rank} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-1]}) \\ \mathcal{H}_1(u_{[0,t-1]}) \end{bmatrix} \tag{7}$$

for all $t = 1, 2, \ldots, T-1$. This increase in rank obviously occurs for any input $u(t) \in \mathbb{R}^m$ if $x(t) \notin \text{im } \mathcal{H}_1(x_{[0,t-1]})$. Therefore, $u(t)$ can be chosen arbitrarily in this case. However, it is not evident that (7) can be guaranteed if $x(t) \in \text{im } \mathcal{H}_1(x_{[0,t-1]})$. Here, the important intermediate step of Theorem 1 is to show that there exists[2] a ξ and a nonzero η satisfying (5). The existence of these vectors heavily relies on controllability of (A, B), and can be proven using a geometric argument that draws some inspiration from Hautus' proof of Heymann's lemma [31].

Once the existence of ξ and $\eta \neq 0$ has been established, the construction of the input $u(t)$ should come without surprise. Indeed, choosing $u(t)$ such that $\xi^\top x(t) + \eta^\top u(t) \neq 0$ ensures that $\begin{bmatrix} \xi^\top & \eta^\top \end{bmatrix}$ is not in the left kernel of the input/state Hankel matrix up to time t. By (5), this implies that

$$\dim \text{leftker} \begin{bmatrix} \mathcal{H}_1(x_{[0,t]}) \\ \mathcal{H}_1(u_{[0,t]}) \end{bmatrix} < \dim \text{leftker} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-1]}) \\ \mathcal{H}_1(u_{[0,t-1]}) \end{bmatrix},$$

which is equivalent to (7).

[2]These vectors can, e.g., be computed using the singular value decomposition of the input/state Hankel matrix.

Note that to guarantee the rank condition (6), we require at least $T \geq n + m$ samples. A surprising fact is that Theorem 1 always guarantees (6) with *exactly* $n + m$ samples, despite the a priori lack of knowledge of the system matrices A and B. This makes our design method completely *sample efficient* in the sense that any experiment design method requires at least as many samples as the one in Theorem 1. In particular, our method outperforms the usual condition of persistency of excitation of order $n+1$, which requires at least $T \geq nm+n+m$ samples to guarantee (6). Specifically, Theorem 1 saves at least nm samples compared to persistency of excitation. The number of samples in Theorem 1 is *linear* (instead of quadratic) in the dimensions of the system variables.

We emphasize that the success of Theorem 1 is due to its *online* nature: the input $u(t)$ is computed by making use of past inputs and states, which contain valuable information about the unknown system. The method is thereby fundamentally different from the classical persistency of excitation condition, which is a purely *offline* condition: one can design a persistently exciting input before collecting any data. In the language of Sontag [32], persistently exciting inputs (of order $n + 1$) are *universal* in the sense that they guarantee (6) *for any* controllable system (1a). In contrast, the input sequence of Theorem 1 is tailored to the *specific* system (1a) that has produced the past inputs and states. This allows for a reduction of the required number of data samples, at the small cost of some simple online computations.

Remark 1: Additional constraints can be considered on the input $u(t)$ in Theorem 1. For example, one can incorporate the norm constraint $\|u(t)\| = \delta$ by selecting

$$u(t) = \begin{cases} \frac{\delta\eta}{\|\eta\|}, & \text{if } \xi^\top x(t) \geq 0 \\ -\frac{\delta\eta}{\|\eta\|}, & \text{otherwise}, \end{cases}$$

since this $u(t)$ satisfies $\xi^\top x(t) + \eta^\top u(t) \neq 0$. Theorem 1 also demonstrates that a sequence of independent and normally distributed random inputs is sample efficient with probability 1. This is because the set $\mathcal{S} = \{u(t) \mid \xi^\top x(t) + \eta^\top u(t) = 0\}$ is a proper affine subspace of $\mathbb{R}^m$. However, we emphasize that Theorem 1 deals with a general class of inputs that yields absolute (rather than probabilistic) guarantees.

B. Input Design for Input/Output Systems

Next, we turn our attention to input/output data generated by system (1). In this setting, we want to find a sequence of inputs $u(0), u(1), \ldots, u(T-1)$ so that the matrix (3) has rank $n+mL$. As before, we opt for the *online* design of these inputs, meaning that we select $u(t)$ on the basis of the collected outputs $y(0), y(1), \ldots, y(t-1)$ and inputs $u(0), u(1), \ldots, u(t-1)$.

The following theorem is a building block in our approach, and shows how full row rank of the *input/state* Hankel matrix

$$\begin{bmatrix} \mathcal{H}_1(x_{[0,T-L]}) \\ \mathcal{H}_L(u_{[0,T-1]}) \end{bmatrix} = \begin{bmatrix} x(0) & x(1) & \cdots & x(T-L) \\ u(0) & u(1) & \cdots & u(T-L) \\ \vdots & \vdots & & \vdots \\ u(L-1) & u(L) & \cdots & u(T-1) \end{bmatrix} \tag{8}$$

can be guaranteed if the state of (1a) is measured.

Theorem 2: Consider the controllable system (1a). Define $T := n + (m+1)L - 1$ for $L \geq 1$. Choose

$u(0), u(1), \ldots, u(L-1) \in \mathbb{R}^m$ arbitrarily, but not all zero. Design $u(t)$ (for $t = L, L+1, \ldots, T-1$) as follows:

- If

$$\begin{bmatrix} x(t-L+1) \\ u_{[t-L+1,t-1]} \end{bmatrix} \notin \mathrm{im} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_{L-1}(u_{[0,t-2]}) \end{bmatrix}, \quad (9)$$

then choose $u(t) \in \mathbb{R}^m$ arbitrarily.

- If

$$\begin{bmatrix} x(t-L+1) \\ u_{[t-L+1,t-1]} \end{bmatrix} \in \mathrm{im} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_{L-1}(u_{[0,t-2]}) \end{bmatrix}, \quad (10)$$

then there exist $\xi \in \mathbb{R}^n$ and $\eta_1, \eta_2, \ldots, \eta_L \in \mathbb{R}^m$ with $\eta_1 \neq 0$ such that

$$\begin{bmatrix} \xi^\top & \eta_L^\top & \cdots & \eta_2^\top & \eta_1^\top \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = 0. \quad (11)$$

In this case, choose $u(t)$ such that

$$\xi^\top x(t-L+1) + \eta_L^\top u(t-L+1) + \cdots + \eta_1^\top u(t) \neq 0.$$

Then, the sequence $u(0), u(1), \ldots, u(T-1)$ is such that

$$\mathrm{rank} \begin{bmatrix} \mathcal{H}_1(x_{[0,T-L]}) \\ \mathcal{H}_L(u_{[0,T-1]}) \end{bmatrix} = n + mL. \quad (12)$$

Proof: Since $u(0), u(1), \ldots, u(L-1)$ are not all zero, we have

$$\mathrm{rank} \begin{bmatrix} x(0) \\ u_{[0,L-1]} \end{bmatrix} = 1.$$

The idea of the proof is to show that in each time step, we can increase the rank of the Hankel matrix. That is, for each $t = L, L+1, \ldots, T-1$, we want to prove that

$$\mathrm{rank} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L+1]}) \\ \mathcal{H}_L(u_{[0,t]}) \end{bmatrix} > \mathrm{rank} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix}. \quad (13)$$

Let $t \in \{L, L+1, \ldots, T-1\}$. First consider the case that (9) holds. Clearly, (13) is satisfied for any $u(t) \in \mathbb{R}^m$. Next, consider the case that (10) holds. We will prove by contradiction that there exist vectors ξ and $\eta_1, \eta_2, \ldots, \eta_L$ with $\eta_1 \neq 0$ satisfying (11). If such vectors do not exist then

$$\mathrm{leftker} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = \mathrm{leftker} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_{L-1}(u_{[0,t-2]}) \end{bmatrix} \times \{0_m\}. \quad (14)$$

We claim that this implies that

$$\mathrm{leftker} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = \mathrm{leftker} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_{L-i}(u_{[0,t-i-1]}) \end{bmatrix} \times \{0_{im}\} \quad (15)$$

for all $i = 1, \ldots, L$. By hypothesis, (15) holds for $i = 1$. Suppose that (15) holds for some $1 \le i = k < L$. Our goal is to show that (15) holds for $i = k+1$ as well. Let

$$\begin{bmatrix} \xi^\top & \eta_L^\top & \cdots & \eta_1^\top \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = 0, \quad (16)$$

where $\xi \in \mathbb{R}^n$ and $\eta_1, \ldots, \eta_L \in \mathbb{R}^m$. By the induction hypothesis, we have $\eta_1 = \cdots = \eta_k = 0$ so that

$$\begin{bmatrix} \xi^\top & \eta_L^\top & \cdots & \eta_{k+1}^\top \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_{L-k}(u_{[0,t-k-1]}) \end{bmatrix} = 0. \quad (17)$$

By (10), we obtain

$$\begin{bmatrix} x(t-L+1) \\ u_{[t-L+1,t-k]} \end{bmatrix} \in \mathrm{im} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_{L-k}(u_{[0,t-k-1]}) \end{bmatrix}.$$

Combining the latter inclusion with (17) yields

$$\begin{bmatrix} \xi^\top & \eta_L^\top & \cdots & \eta_{k+1}^\top \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L+1]}) \\ \mathcal{H}_{L-k}(u_{[0,t-k]}) \end{bmatrix} = 0. \quad (18)$$

Note that the laws of system (1a) imply that

$$\begin{bmatrix} \mathcal{H}_1(x_{[1,t-L+1]}) \\ \mathcal{H}_{L-k}(u_{[1,t-k]}) \end{bmatrix} = \begin{bmatrix} A & B & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_{L-k+1}(u_{[0,t-k]}) \end{bmatrix}.$$

Therefore, (18) implies

$$\begin{bmatrix} \xi^\top A & \xi^\top B & \eta_L^\top & \cdots & \eta_{k+1}^\top \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_{L-k+1}(u_{[0,t-k]}) \end{bmatrix} = 0.$$

This yields

$$\begin{bmatrix} \xi^\top A & \xi^\top B & \eta_L^\top & \cdots & \eta_{k+1}^\top & 0_{(k-1)m}^\top \end{bmatrix} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = 0.$$

Finally, by the induction hypothesis we see that $\eta_{k+1} = 0$, as desired. Thus, (15) holds for all $i = 1, \ldots, L$. In particular, for $i = L$ we obtain

$$\mathrm{leftker} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = \mathrm{leftker} \, \mathcal{H}_1(x_{[0,t-L]}) \times \{0_{mL}\},$$

equivalently (using the fact that $(\mathrm{im}\, X)^\perp = \mathrm{leftker}\, X$),

$$\mathrm{im} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = \mathrm{im} \, \mathcal{H}_1(x_{[0,t-L]}) \times \mathbb{R}^{mL}. \quad (19)$$

Multiplication on both sides by the matrix $\begin{bmatrix} A & B & 0 \end{bmatrix}$ yields

$$A \, \mathrm{im} \, \mathcal{H}_1(x_{[0,t-L]}) + \mathrm{im} \, B = \mathrm{im} \, \mathcal{H}_1(x_{[1,t-L+1]}).$$

Using (10), we obtain

$$A \, \mathrm{im} \, \mathcal{H}_1(x_{[0,t-L]}) + \mathrm{im} \, B \subseteq \mathrm{im} \, \mathcal{H}_1(x_{[0,t-L]}).$$

In particular, we see that

$$A \, \mathrm{im} \, \mathcal{H}_1(x_{[0,t-L]}) \subseteq \mathrm{im} \, \mathcal{H}_1(x_{[0,t-L]})$$
$$\mathrm{im} \, B \subseteq \mathrm{im} \, \mathcal{H}_1(x_{[0,t-L]}).$$

In other words, $\mathrm{im} \, \mathcal{H}_1(x_{[0,t-L]})$ is an A-invariant subspace containing $\mathrm{im}\, B$. Since the reachable subspace $\langle A \mid \mathrm{im}\, B \rangle$ of the pair (A, B) is the *smallest* A-invariant subspace containing $\mathrm{im}\, B$ (see [33, Ch. 3]), we see that

$$\mathbb{R}^n = \langle A \mid \mathrm{im}\, B \rangle \subseteq \mathrm{im} \, \mathcal{H}_1(x_{[0,t-L]}),$$

where we made use of the fact that (A, B) is controllable. As such, $\mathrm{im} \, \mathcal{H}_1(x_{[0,t-L]}) = \mathbb{R}^n$ and by (19) we see that

$$\mathrm{im} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = \mathbb{R}^{n+mL}.$$

This implies that $t \ge n + (m+1)L - 1$ which leads to a contradiction as $t \in \{L, L+1, \ldots, T-1\}$. As such, we conclude that (14) does not hold. Therefore, there exist $\xi \in \mathbb{R}^n$ and $\eta_1, \eta_2, \ldots, \eta_L \in \mathbb{R}^m$ with $\eta_1 \neq 0$ such that (11) holds. Clearly, this means that there exists a $u(t)$ such that

$$\xi^\top x(t-L+1) + \eta_L^\top u(t-L+1) + \cdots + \eta_1^\top u(t) \neq 0.$$

For such an input, we have that

$$\dim \text{leftker} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} > \dim \text{leftker} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L+1]}) \\ \mathcal{H}_L(u_{[0,t]}) \end{bmatrix},$$

and consequently, (13) holds. This proves the theorem. ∎

Note that in the special case $L = 1$, Theorem 2 reduces to Theorem 1. Next, we turn our attention to the situation in which we measure inputs and outputs and want to ensure that (3) has rank $n + mL$. This is the topic of the next theorem.

Theorem 3: Consider the minimal system (1). Let $L > \ell$ and $T := n + (m+1)L - 1$. Select $u(0), u(1), \ldots, u(L-1) \in \mathbb{R}^m$ arbitrarily, but not all zero. Furthermore, design $u(t)$ (for $t = L, L+1, \ldots, T-1$) as follows:

- If
$$\begin{bmatrix} y_{[t-L+1,t-1]} \\ u_{[t-L+1,t-1]} \end{bmatrix} \notin \text{im} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,t-2]}) \\ \mathcal{H}_{L-1}(u_{[0,t-2]}) \end{bmatrix}, \quad (20)$$
then choose $u(t) \in \mathbb{R}^m$ arbitrarily.

- If
$$\begin{bmatrix} y_{[t-L+1,t-1]} \\ u_{[t-L+1,t-1]} \end{bmatrix} \in \text{im} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,t-2]}) \\ \mathcal{H}_{L-1}(u_{[0,t-2]}) \end{bmatrix}, \quad (21)$$
there exist $\xi_1, \ldots, \xi_{L-1} \in \mathbb{R}^p$ and $\eta_1, \ldots, \eta_L \in \mathbb{R}^m$ with $\eta_1 \neq 0$ such that
$$\begin{bmatrix} \xi_{L-1}^\top \cdots \xi_1^\top \eta_L^\top \cdots \eta_1^\top \end{bmatrix} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,t-2]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = 0. \quad (22)$$
In this case, choose $u(t)$ such that
$$\xi_{L-1}^\top y(t-L+1) + \cdots + \xi_1^\top y(t-1)$$
$$+ \eta_L^\top u(t-L+1) + \cdots + \eta_1^\top u(t) \neq 0. \quad (23)$$

Then we have that
$$\text{rank} \begin{bmatrix} \mathcal{H}_L(y_{[0,T-1]}) \\ \mathcal{H}_L(u_{[0,T-1]}) \end{bmatrix} = n + mL. \quad (24)$$

Proof: Since $u(0), u(1), \ldots, u(L-1)$ are not all zero, we have
$$\text{rank} \begin{bmatrix} y_{[0,L-2]} \\ u_{[0,L-1]} \end{bmatrix} = 1.$$

The idea of the proof is to show that for each time step $t = L, L+1, \ldots, T-1$, we have
$$\text{rank} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,t-1]}) \\ \mathcal{H}_L(u_{[0,t]}) \end{bmatrix} > \text{rank} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,t-2]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix}. \quad (25)$$

This would prove that
$$\text{rank} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,T-2]}) \\ \mathcal{H}_L(u_{[0,T-1]}) \end{bmatrix} = n + mL,$$

and consequently, (24) holds.

Let $t \in \{L, L+1, \ldots, T-1\}$. First consider the case that (20) holds. Clearly, (25) is satisfied for any $u(t) \in \mathbb{R}^m$.

Next, consider the case that (21) holds. We claim that
$$\text{leftker} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,t-2]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} \neq \text{leftker} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,t-2]}) \\ \mathcal{H}_{L-1}(u_{[0,t-2]}) \end{bmatrix} \times \{0_m\}. \quad (26)$$

We will prove this claim by contradiction. Thus, suppose that (26) does hold with equality, equivalently,
$$\text{im} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,t-2]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = \text{im} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,t-2]}) \\ \mathcal{H}_{L-1}(u_{[0,t-2]}) \end{bmatrix} \times \mathbb{R}^m. \quad (27)$$

Define the matrices N and M as
$$N := \begin{bmatrix} \mathcal{O}_{L-1} & \mathcal{T}_{L-1} \\ 0 & I \end{bmatrix} \text{ and } M := \begin{bmatrix} N & 0 \\ 0 & I_m \end{bmatrix},$$

where the observability matrix $\mathcal{O}_i$ is defined in (2) and the Toeplitz matrix $\mathcal{T}_i$ is given by
$$\mathcal{T}_i := \begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \cdots & D \end{bmatrix}.$$

Then (27) implies
$$M \text{ im} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = M \left(\text{im} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_{L-1}(u_{[0,t-2]}) \end{bmatrix} \times \mathbb{R}^m \right).$$

Since (C, A) is observable and $L > \ell$, the matrix M has full column rank. As such, we conclude that
$$\text{im} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} = \text{im} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_{L-1}(u_{[0,t-2]}) \end{bmatrix} \times \mathbb{R}^m,$$

i.e., (14) holds. Similarly, we note that (21) implies
$$N \begin{bmatrix} x(t-L+1) \\ u_{[t-L+1,t-1]} \end{bmatrix} \in N \text{ im} \begin{bmatrix} \mathcal{H}_1(x_{[0,t-L]}) \\ \mathcal{H}_{L-1}(u_{[0,t-2]}) \end{bmatrix}.$$

Since N has full column rank, this implies that (10) holds. It now follows from the proof of Theorem 2 that t satisfies $t \geq n + (m+1)L - 1$. This yields a contradiction since $t \in \{L, L+1, \ldots, T-1\}$. As such, (26) holds. Therefore, there exist $\xi_1, \xi_2, \ldots, \xi_{L-1} \in \mathbb{R}^p$ and $\eta_1, \eta_2, \ldots, \eta_L \in \mathbb{R}^m$ with $\eta_1 \neq 0$ such that (22) is satisfied. This means that we can choose an input $u(t)$ such that (23) holds. For this choice of $u(t)$, we obtain
$$\dim \text{leftker} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,t-2]}) \\ \mathcal{H}_L(u_{[0,t-1]}) \end{bmatrix} > \dim \text{leftker} \begin{bmatrix} \mathcal{H}_{L-1}(y_{[0,t-1]}) \\ \mathcal{H}_L(u_{[0,t]}) \end{bmatrix}.$$

We conclude that (25) holds, which proves the theorem. ∎

We note that (24) can only hold if $T \geq n + (m+1)L - 1$. By Theorem 3, we can always design an informative experiment of length *exactly* $n + (m+1)L - 1$. In other words, the design procedure of Theorem 3 is again sample efficient. In comparison, the usual persistency of excitation condition of order $n+L$ (Proposition 1) requires $T \geq n + (m+1)L + mn - 1$ samples. Theorem 3 thus saves at least mn samples when compared to persistency of excitation.

Finally, is it noteworthy that exact knowledge of the state-space dimension is *not required* for Theorem 3, even though the time horizon T is defined in terms of n. The only requirement is an upper bound $L > \ell$ on the lag of the system. The reason is that all the steps of Theorem 3 can be executed without knowledge of n. One should only take some care with the stopping criterion since T is a priori unknown. A simple approach to deal with this is the following: apply the steps of Theorem 3 for $t = 0, 1, \ldots,$ until t is such that (27) is satisfied. By the proof of Theorem 3, the smallest t for which (27) holds is $t = n + (m+1)L - 1$. From this relation, n can be recovered, and for $T := t$, the rank condition (24) holds.

IV. CONCLUSION AND DISCUSSION

The purpose of this letter has been to provide a new angle of attack to the problem of input design for data-driven modeling and control of linear systems. Instead of the classical persistency of excitation condition, we have proposed an *online* method to select the inputs. We have studied both input/state and input/output systems. For both types of systems, we have provided an input design method that guarantees an important rank condition on the data Hankel matrix. A notable feature of our approach is that it is completely *sample efficient* in the sense that it guarantees these rank conditions with the minimum number of samples.

The presented results apply to exact data, and the generalization to noisy data is arguably the most important problem for future work. In the case of noisy measurements, rank conditions on Hankel matrices are not sufficient to perform successful data-driven modeling and control. Nonetheless, under suitable conditions, controllers with guaranteed stability and performance can be obtained from data. For this, data-based linear matrix inequality (LMI) conditions were studied in [4] and [6], and a necessary and sufficient condition was provided in [7] using a generalization of Yakubovich's S-lemma [34].

Future work could thus focus on experiment design techniques that aim at making such data-based LMI's feasible. One possible avenue to explore is to apply the repeated dimension reduction argument in this letter to other sets. In the current work, we have shown that it is possible to reduce the dimension of the *left kernel* of a Hankel matrix at every time step. However, in the above context of data-driven control, a fundamental role is played by the solution set to the *dual LMI* [35] of the data-based LMI's for control. This dual LMI is infeasible if and only if its primal LMI is feasible. Thus, a systematic reduction of the dimension of the dual solution set would guarantee feasibility of the data-based LMI's in a finite number of steps. Constructing inputs that achieve this is a main challenge for future research.

REFERENCES

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

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

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

[4] 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, Mar. 2020.

[5] 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, Nov. 2020.

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

[7] 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*, early access, Dec. 25, 2020, doi: 10.1109/TAC.2020.3047577.

[8] J. Coulson, J. Lygeros, and F. Dörfler, "Distributionally robust chance constrained data-enabled predictive control," 2020. [Online]. Available: arxiv.org/abs/2006.01702.

[9] A. Xue and N. Matni, "Data-driven system level synthesis," 2020. [Online]. Available: arxiv.org/abs/2011.10674.

[10] H. J. van Waarde and M. Mesbahi, "Data-driven parameterizations of suboptimal LQR and H2 controllers," 2020. [Online]. Available: arxiv.org/abs/1912.07671.

[11] M. Guo, C. De Persis, and P. Tesi, "Learning control for polynomial systems using sum of squares relaxations," in *Proc. 59th IEEE Conf. Decis. Control*, Jeju, South Korea, 2020, pp. 2436–2441.

[12] T. Dai and M. Sznaier, "A semi-algebraic optimization approach to data-driven control of continuous-time nonlinear systems," *IEEE Control Syst. Lett.*, vol. 5, no. 2, pp. 487–492, Apr. 2021.

[13] G. C. Goodwin and R. L. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*. New York, NY, USA: Academic, 1977.

[14] M. Gevers and L. Ljung, "Optimal experiment designs with respect to the intended model application," *Automatica*, vol. 22, no. 5, pp. 543–554, 1986.

[15] H. Jansson and H. Hjalmarsson, "Input design via LMIs admitting frequency-wise model specifications in confidence regions," *IEEE Trans. Autom. Control*, vol. 50, no. 10, pp. 1534–1549, Oct. 2005.

[16] X. Bombois, G. Scorletti, M. Gevers, P. M. J. Van den Hof, and R. Hildebrand, "Least costly identification experiment for control," *Automatica*, vol. 42, no. 10, pp. 1651–1662, 2006.

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

[18] F. Fabiani and P. J. Goulart, "The optimal transport paradigm enables data compression in data-driven robust control," 2020. [Online]. Available: arxiv.org/abs/2005.09393.

[19] M. Yin, A. Iannelli, and R. S. Smith, "Maximum likelihood estimation in data-driven modeling and control," 2020. [Online]. Available: arxiv.org/abs/2011.00925.

[20] A. Iannelli, M. Yin, and R. S. Smith, "Experiment design for impulse response identification with signal matrix models," 2020. [Online]. Available: arxiv.org/abs/2012.08126.

[21] H. Hjalmarsson, M. Gevers, S. Gunnarsson, and O. Lequin, "Iterative feedback tuning: Theory and applications," *IEEE Control Syst. Mag.*, vol. 18, no. 4, pp. 26–41, Aug. 1998.

[22] N. Filatov and H. Unbehauen, *Adaptive Dual Control: Theory and Applications* (Lecture Notes in Control and Information Sciences). Heidelberg, Germany: Springer, 2004.

[23] G. Marafioti, R. R. Bitmead, and M. Hovd, "Persistently exciting model predictive control," *Int. J. Adapt. Control Signal Process.*, vol. 28, no. 6, pp. 536–552, 2014.

[24] C. A. Larsson, A. Ebadat, C. R. Rojas, X. Bombois, and H. Hjalmarsson, "An application-oriented approach to dual control with excitation for closed-loop identification," *Eur. J. Control*, vol. 29, pp. 1–16, May 2016.

[25] K. Lindqvist and H. Hjalmarsson, "Identification for control: Adaptive input design using convex optimization," in *Proc. 40th IEEE Conf. Decis. Control*, vol. 5. Orlando, FL, USA, 2001, pp. 4326–4331.

[26] L. Gerencsér, H. Hjalmarsson, and L. Huang, "Adaptive input design for LTI systems," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2390–2405, May 2017.

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

[28] I. Markovsky and F. Dörfler. (2020). *Identifiability in the Behavioral Setting*. [Online]. Available: home-pages.vub.ac.be/imarkovs/publications/identifiability.pdf

[29] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[30] 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, Jul. 2020.

[31] M. Hautus, "A simple proof of Heymann's lemma," *IEEE Trans. Autom. Control*, vol. 22, no. 5, pp. 885–886, Oct. 1977.

[32] E. Sontag, "On the length of inputs necessary in order to identify a deterministic linear system," *IEEE Trans. Autom. Control*, vol. 25, no. 1, pp. 120–121, Feb. 1980.

[33] H. L. Trentelman, A. A. Stoorvogel, and M. Hautus, *Control Theory for Linear Systems*. London, U.K.: Springer-Verlag, 2001.

[34] V. A. Yakubovich, "S-procedure in nonlinear control theory," *Vestnik Leningrad Univ. Math.*, vol. 4, no. 1, pp. 73–93, 1977.

[35] V. Balakrishnan and L. Vandenberghe, "Semidefinite programming duality and linear time-invariant systems," *IEEE Trans. Autom. Control*, vol. 48, no. 1, pp. 30–41, Jan. 2003.