On the design of persistently exciting inputs for data-driven control of linear and nonlinear systems

Mohammad Alsalti, Victor G. Lopez, and Matthias A. Müller

Abstract— In the context of data-driven control, persistence of excitation (PE) of an input sequence is defined in terms of a rank condition on the Hankel matrix of the input data. For nonlinear systems, recent results employed rank conditions involving collected input and state/output data, for which no guidelines are available on how to satisfy them a priori. In this paper, we first show that a set of discrete impulses is guaranteed to be persistently exciting for any controllable LTI system. Based on this result, for certain classes of nonlinear systems, we guarantee persistence of excitation of sequences of basis functions *a priori*, by design of the physical input only.

I. INTRODUCTION

Central to the area of data-driven control, as well as system identification [1] and adaptive control [2], is the notion of persistence of excitation (PE), see, e.g., [3]. In [4], a discrete-time sequence is said to be persistently exciting of a certain order if a corresponding Hankel matrix of that sequence has full row rank. It was then shown in [4] that when such a PE input is applied to a controllable LTI system, the resulting input-state or input-output Hankel matrices satisfy certain rank conditions. Now known as the fundamental lemma, this result has received a great amount of attention in recent years and has been successfully used in a wide range of applications (see [5] for a comprehensive survey).

For certain classes of nonlinear systems, various works have proposed extensions and applications of the fundamental lemma using, e.g., basis functions. For instance, Hammerstein-Wiener systems [6], flat and feedback linearizable systems [7], [8], and control design of input-affine nonlinear systems [9] have been considered. In these works, suitable PE conditions involving the sequence of basis functions (which depend on inputs and/or states/outputs) need to be satisfied. However, in [6]–[9] these PE conditions could only be verified a posteriori, i.e., after performing an experiment and collecting state/output data, and no a priori input design was proposed to this end.

In general, there exist only few results on the design of suitable inputs that result in satisfaction of the required PE conditions, specifically for nonlinear systems. In [10], an *online* method was proposed to design inputs that result in the desired rank conditions of [4] on input-state or input-output Hankel matrices for linear systems. However, the resulting input is not universal in the sense that it is tailored specifically to the system on which the experiment is performed.

Leibniz University Hannover, Institute of Automatic Control, 30167 Hannover, Germany. E-mail:{alsalti,lopez,mueller}@irt.uni-hannover.de.

An extension of this result to the class of bilinear systems appeared in [11]. Furthermore, [12] shows how suitable scaling of the initial conditions and the input to a nonlinear system leads to satisfaction of certain rank conditions on the input-state data. This enables local stabilization of an unknown nonlinear system in the first approximation but, in general, cannot be used for the methods in [6]–[9], [13].

The contributions of this paper are as follows. First, we propose a simple input sequence which satisfies standard PE conditions for linear systems [3], [4]. Based on this result, as a second contribution, we design inputs that guarantee PE of any order for an arbitrary sequence of basis functions for the class of Hammerstein systems. The third contribution addresses nonlinear systems that are locally reachable at the origin. Specifically, we show existence of sparse input sequences that guarantee collective PE of sequences of basis functions. For single-input single-output (SISO) flat systems (which are locally reachable at the origin), we systematically design inputs that guarantee collective PE of any order of specific choices of basis functions. Finally, we illustrate the results by computing data-based controllers for SISO flat systems (as proposed in [9]).

Section II introduces notation and necessary preliminaries. Sections III-V contain the main contributions of the paper. Section VI illustrates the results with a numerical example and Section VII concludes the paper.

II. NOTATION AND PRELIMINARIES

Notation: Let $\mathbb{Z}_{>0}$ denote the set of positive integers and let $\mathbb{Z}_{[a,b]}$ denote the set of integers in the interval $[a,b]$. Let the $m \times m$ identity matrix be I_m and its columns be e_i for $i \in \mathbb{Z}_{[1,m]}$. We use $\mathbf{0}_{n \times m}$ to denote an $n \times m$ matrix of zeros; when the dimensions are clear, the subscript is omitted. For a sequence $\{z_k\}_{k=0}^{N-1}$ with $z_k \in \mathbb{R}^\eta$, we denote its stacked vector as $z = \begin{bmatrix} z_0^\top & z_1^\top & \dots & z_{N-1}^\top \end{bmatrix}^\top$ and a stacked window of it as $z_{[l,j]} = \begin{bmatrix} z_l^\top & z_{l+1}^\top & \dots & z_j^\top \end{bmatrix}^\top$ for $0 \leq l < j$. We write $M \succ \mathbf{0}$ ($M \succeq \mathbf{0}$) if the matrix M is positive (semi-)definite.

In this paper, we are concerned with the notion of persistence of excitation which originated in the fields of system identification [1] and adaptive control [2]. Several definitions of PE have appeared in the literature. The following is one of such definitions for a finite length discrete-time sequence.

Definition 1 ([3]): For $N \in \mathbb{Z}_{>0}$, the sequence $\{z_k\}_{k=0}^{N-1}$, $z_k \in \mathbb{R}^\eta$, is exciting over the interval $[0, N-1]$ if, for some $\nu > 0$, the following holds $\sum_{k=0}^{N-1} z_k z_k^\top \succeq \nu I_\eta \succ \mathbf{0}$.

In recent works on data-driven control (cf. [5]), the following definition has been commonly used.

Definition 2 ([4]): For $L \in \mathbb{Z}_{>0}$ and $N \geq L$, the sequence $\{z_k\}_{k=0}^{N-1}$, $z_k \in \mathbb{R}^\eta$, is persistently exciting of order L if $\mathrm{rank}(\mathscr{H}_L(z)) = \eta L$, where $\mathscr{H}_L(z) = \begin{bmatrix} z_{[0,L-1]} & z_{[1,L]} & \cdots & z_{[N-L,N-1]} \end{bmatrix}$.

The two definitions are related as follows: Notice that if a sequence is PE of order L in the sense of Definition 2, then

$$\mathscr{H}_L(z)(\mathscr{H}_L(z))^\top = \sum_{k=0}^{N-L} z_{[k,k+L-1]} z_{[k,k+L-1]}^\top \succ \mathbf{0}.$$

This means that for $L = 1$, the two definitions are equivalent. If a sequence is PE in the sense of Definition 2 of order $L > 1$, then it is also exciting in the sense of Definition 1, but the converse is not necessarily true. This is because if $\mathrm{rank}(\mathscr{H}_L(z)) = \eta L$, then $\mathrm{rank}(\mathscr{H}_{\bar{L}}(z)) = \eta \bar{L}$ for any depth $\bar{L} < L$, but the converse is, in general, not true.

The advantage of Definition 2 is that it quantifies the order of which a signal is exciting. As shown in [4], this notion has an important application. In particular, if an input to a controllable LTI system is PE of order $L + n$, then the resulting input/output data matrix contains in its span any length$-L$ input/output trajectory of the system. This became known as the fundamental lemma and is summarized below.

Theorem 1 ([4]): Let $\{u_k, y_k\}_{k=0}^{N-1}$ be an input-output trajectory of a controllable LTI system. If $\{u_k\}_{k=0}^{N-1}$ is PE of order $L + n$, then any $\{\bar{u}_k, \bar{y}_k\}_{k=0}^{L-1}$ is a trajectory of the system, if and only if there exists $\beta \in \mathbb{R}^{N-L+1}$ such that

$$\begin{bmatrix} \mathscr{H}_L(u) \\ \mathscr{H}_L(y) \end{bmatrix} \beta = \begin{bmatrix} \bar{u} \\ \bar{y} \end{bmatrix}. \tag{1}$$

The following are recent extensions of Definition 2, which we use in our paper. The notion of *collective* persistence of excitation was defined in [14], and extends Definition 2 to multiple sequences.

Definition 3 ([14]): For $r, L \in \mathbb{Z}_{>0}$, $j \in \mathbb{Z}_{[1,r]}$, and $N_j \geq L$, the sequences $\{z_k^{(j)}\}_{k=0}^{N_j-1}$, with $z_k^{(j)} \in \mathbb{R}^\eta$, are *collectively persistently exciting* of order L if $\mathrm{rank}(\mathcal{H}_L(\mathscr{Z})) = \eta L$, where $\mathscr{Z} = \begin{bmatrix} (z^{(1)})^\top & \cdots & (z^{(r)})^\top \end{bmatrix}^\top$, and

$$\mathcal{H}_L(\mathscr{Z}) = \begin{bmatrix} \mathscr{H}_L(z^{(1)}) & \cdots & \mathscr{H}_L(z^{(r)}) \end{bmatrix}.$$

Another extension was proposed in [15], where a *quantitative* notion of PE is defined in order to establish a robust version of Theorem 1. There, a sequence $\{z_k\}_{k=0}^{N-1}$ is said to be $\alpha-$PE of order L if $\sigma_{\min}(\mathscr{H}_L(z)) \geq \alpha > 0$, provided the Hankel matrix has at least as many columns as rows, where $\sigma_{\min}$ denotes the minimum singular value.

In the next section, we propose a simple input sequence which satisfies all the above definitions of PE.

III. PE FOR LTI SYSTEMS

It is well known that random inputs satisfy Definition 2 with high probability. However, this is not guaranteed and in some practical applications it may not be possible/desirable to apply frequent inputs. This is the case, e.g., in networked systems with bandwidth constraints [16] or closed-loop control of fluid resuscitation [17], where sparse inputs in the form of boluses are used to identify patient-specific parameters.

In the following theorem, we propose a simple and highly sparse input sequence that is guaranteed to be PE. In particular, PE of order $L \in \mathbb{Z}_{>0}$ is achieved by giving a pulse at each of the m input channels every L instants.

Theorem 2: Let $L \in \mathbb{Z}_{>0}$ and $N \geq (m+1)L - 1$. If the sequence $\{u_k\}_{k=0}^{N-1}$ takes the form

$$u_k = \begin{cases} e_j, & k = jL - 1, \\ \mathbf{0}, & \text{otherwise,} \end{cases} \tag{2}$$

where $j \in \mathbb{Z}_{[1,m]}$, then it holds that $\mathrm{rank}(\mathscr{H}_L(u)) = mL$.

Proof: Without loss of generality, let $N = (m+1)L - 1$. The following matrix has full row rank

$$\mathscr{H}_L(u) = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & e_1 & \cdots & \mathbf{0} & \mathbf{0} & \cdots & e_m \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & e_1 & \cdots & \mathbf{0} & \cdots & \mathbf{0} & e_m & \cdots & \mathbf{0} \\ e_1 & \mathbf{0} & \cdots & \mathbf{0} & \cdots & e_m & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}. \quad \blacksquare$$

Remark 1: For the input in (2), all singular values of $\mathscr{H}_L(u)$ are equal to one. By scaling (2) by α, one obtains an input which is $\alpha-$PE of order L (cf. [15]). Note that larger values of α imply higher levels of PE of order L.

Theorem 2 provides an explicit formula (2) for an input which is guaranteed to be persistently exciting in the sense of Definition 2 (and hence, also Definition 1). This can be used to apply the results of Theorem 1. A necessary condition for the input in (2) to be PE of order $L + n$ is that it is at least of length $N \geq (m+1)(L+n) - 1$ (cf. Definition 2). In contrast, the procedure from [10] uses online state or output measurements to design an input such that the resulting input-output data matrix in (1) has rank $mL + n$ and requires $N = (m+1)L + n - 1$ samples to this end, implying that it is sample efficient. However, the resulting input only guarantees the rank condition for the system on which the experiment was done. In contrast, the input (2), although not as sample efficient, guarantees the results of Theorem 1 independently of the considered system.

Apart from its use in data-driven analysis and control, Theorem 1 also allows for identification of the LTI system's matrices up to a similarity transformation (cf. [4, Sec. 4.4]). In this sense, employing (2) to obtain (1) can be interpreted as the data-driven counterpart to classical works on identification of LTI systems from their impulse response [18].

In Section IV, we consider Hammerstein nonlinear systems and show how to guarantee PE of basis functions that depend on the input, by the design of the input only.

IV. PE FOR HAMMERSTEIN SYSTEMS

Consider an unknown Hammerstein system of the form

$$x_{k+1} = Ax_k + B\gamma(u_k), \quad y_k = Cx_k + D\gamma(u_k), \tag{3}$$

with (A, B) controllable, $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, and $y_k \in \mathbb{R}^p$. Furthermore, $\gamma : \mathbb{R}^m \to \mathbb{R}^{\bar{m}}$ is an unknown nonlinear function which satisfies $\gamma(\mathbf{0}) = \mathbf{0}$ without loss of generality. An extension of Theorem 1 to the class of Hammerstein systems appeared in [6], assuming that γ_i, $i \in \mathbb{Z}_{[1,\bar{m}]}$, belong to the span of a given set of r basis functions $\psi_j : \mathbb{R}^m \to \mathbb{R}$, $j \in \mathbb{Z}_{[1,r]}$, which satisfy $\psi_j(\mathbf{0}) = 0$ (note that functions with $\psi_j(\mathbf{0}) \neq 0$ can be suitably shifted by a constant). In this case,

PE must be imposed on the sequence of basis functions. In this section, we show how this can be done only by the design of the physical input u.

We denote the stacked vector of the basis functions by $\Psi(u_k) = \left[\psi_1(u_k)\cdots\psi_r(u_k)\right]^\top$. For $\lambda_j \in \mathbb{R}^m$, $j \in \mathbb{Z}_{[1,r]}$, we define the following square matrix $\Lambda \in \mathbb{R}^{r\times r}$

$$\Lambda = \begin{bmatrix} \Psi(\lambda_1) & \Psi(\lambda_2) & \cdots & \Psi(\lambda_r) \end{bmatrix}. \tag{4}$$

The following theorem provides conditions for an input sequence $\{u_k\}_{k=0}^{N-1}$ that is guaranteed to result in a PE sequence of basis functions $\{\hat{\Psi}_k\}_{k=0}^{N-1}$, where $\hat{\Psi}_k := \Psi(u_k)$.

Theorem 3: Given $L \in \mathbb{Z}_{>0}$ and any r linearly independent basis functions $\psi_j : \mathbb{R}^m \to \mathbb{R}$ which satisfy $\psi_j(\mathbf{0}) = 0$, let $N \geq (r+1)L - 1$. Let $\{u_k\}_{k=0}^{N-1}$ take the form

$$u_k = \begin{cases} \lambda_j, & k = jL - 1, \\ \mathbf{0}, & \text{otherwise}, \end{cases} \tag{5}$$

where $j \in \mathbb{Z}_{[1,r]}$. If $\lambda_j \in \mathbb{R}^m$ are such that the matrix Λ in (4) is invertible, then it holds that $\text{rank}(\mathscr{H}_L(\hat{\Psi})) = rL$.

Proof: Let $\{\tilde{u}_k\}_{k=0}^{N-1}$, $\tilde{u}_k \in \mathbb{R}^r$, take the form of (2). Then by Theorem 2 it holds that $\text{rank}(\mathscr{H}_L(\tilde{u})) = rL$. Next, notice that $\mathscr{H}_L(\hat{\Psi}) = (I_L \otimes \Lambda)\mathscr{H}_L(\tilde{u})$, where $\otimes$ denotes the Kronecker product. Since Λ is invertible by assumption, it follows that $(I_L \otimes \Lambda)$ is a square full rank matrix and hence $\text{rank}(\mathscr{H}_L(\hat{\Psi})) = \text{rank}(\mathscr{H}_L(\tilde{u})) = rL$. ∎

In order to find the values of λ_j which make Λ invertible, one can formulate a feasibility problem as follows

$$\text{find} \quad \lambda_1,\ldots,\lambda_r, \qquad \text{s.t.} \quad \text{rank}(\Lambda) = r. \tag{6}$$

This is equivalent to requiring that $\Lambda^\top\Lambda \succ \mathbf{0}$. Such a problem can be solved offline, e.g., iteratively as in [19] or by reformulating it as a regularized unconstrained nonlinear least squares problem [20]. Solving (6) only requires knowledge of the user-defined basis functions. The following theorem shows that for *any* given set of linearly independent basis functions, (6) is feasible.

Theorem 4: Given any set of r linearly independent basis functions $\psi_j : \mathbb{R}^m \to \mathbb{R}$, there exist $\lambda_j \in \mathbb{R}^m$, for $j \in \mathbb{Z}_{[1,r]}$, such that Λ in (4) is invertible.

Proof: Suppose by contradiction that the maximum rank that Λ can attain for arbitrary $\lambda_j \in \mathbb{R}^m$, $j \in \mathbb{Z}_{[1,r]}$, is $d < r$. Consider such a choice of $\lambda_j \in \mathbb{R}^m$, $j \in \mathbb{Z}_{[1,r]}$, which results in $\text{rank}(\Lambda) = d < r$. Then, there exists a non-zero vector $\rho \in \mathbb{R}^r$ such that $\rho^\top\Lambda = \mathbf{0}$. Moreover, for any $\bar{\lambda} \in \mathbb{R}^m$, the matrix $\begin{bmatrix} \Lambda & \Psi(\bar{\lambda}) \end{bmatrix}$ has the same rank as Λ. Hence, it holds that $\rho^\top\Psi(\bar{\lambda}) = 0$ for arbitrary $\bar{\lambda} \in \mathbb{R}^m$, i.e.,

$$\rho_1\psi_1(\bar{\lambda}) + \cdots + \rho_r\psi_r(\bar{\lambda}) = 0, \qquad \forall\bar{\lambda} \in \mathbb{R}^m. \tag{7}$$

This, however, contradicts linear independence of the basis functions, thus proving that there exists $\lambda_j \in \mathbb{R}^m$, $j \in \mathbb{Z}_{[1,r]}$, such that Λ is invertible. ∎

In the next section, we study the class of locally reachable nonlinear systems at the origin and show existence of sparse inputs that guarantee collective PE of basis functions. For SISO flat systems (which are locally reachable), we show how to design those inputs such that PE of any order $L > 0$ can be guaranteed for a specific choice of basis functions.

V. PE FOR LOCALLY REACHABLE NONLINEAR SYSTEMS

Consider an unknown nonlinear system of the form

$$x_{k+1} = f(x_k, u_k), \tag{8}$$

with $x_k \in \mathbb{R}^n, u_k \in \mathbb{R}^m$ being the state and input vectors, respectively, and $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is an unknown function satisfying $f(\mathbf{0},\mathbf{0}) = \mathbf{0}$. For $\mu \in \mathbb{Z}_{>0}$, the set of all states which can be reached from x_0 in μ steps is defined as

$$\mathcal{R}_\mu(x_0) = \left\{ x_\mu \in \mathbb{R}^n \,\middle|\, \begin{array}{l} \exists\, u_{[0,\mu-1]},\, u_k \in \mathbb{R}^m, \\ \text{s.t. } x_{k+1} = f(x_k, u_k),\, \forall k \in \mathbb{Z}_{[0,\mu-1]}. \end{array} \right\}.$$

It was shown in [21] how one can, under certain assumptions, obtain a guaranteed under-approximation of the reachable set of a nonlinear system with unknown dynamics. For the remainder of this section, we make the following assumption.

Assumption 1: For the system (8), there exists $\mu \in \mathbb{Z}_{>0}$ such that the origin is contained in the interior of $\mathcal{R}_\mu(\mathbf{0})$.

Assumption 1 implies that the system is locally reachable at the origin. A sufficient condition for local reachability at $x = \mathbf{0}$ is that the linearization of system (8) at the origin is controllable (cf. [22, Lemma 3.7.8]).

Consider now r basis functions $\theta_j : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ which satisfy $\theta_j(\mathbf{0},\mathbf{0}) = 0$, $j \in \mathbb{Z}_{[1,r]}$, and denote the stacked vector of them by $\Theta(x_k, u_k) = \left[\theta_1(x_k,u_k)\cdots\theta_r(x_k,u_k)\right]^\top$. Suppose that the functions are linearly independent on arbitrary domains with non-empty interior[1] $D_x \times D_u \subset \mathbb{R}^n \times \mathbb{R}^m$. The objective is to design inputs $\{u_k^{(j)}\}_{k=0}^{N_j-1}$ such that the sequences of basis functions $\{\hat{\Theta}_k^{(j)}\}_{k=0}^{N_j-1}$ (with $\hat{\Theta}_k^{(j)} := \Theta(x_k^{(j)}, u_k^{(j)})$) are collectively persistently exciting of order L. According to Definition 3, the following mosaic Hankel matrix must have full row rank

$$\mathcal{H}_L(\vartheta) = \begin{bmatrix} \mathscr{H}_L(\hat{\Theta}^{(1)}) & \cdots & \mathscr{H}_L(\hat{\Theta}^{(r)}) \end{bmatrix}, \tag{9}$$

where $\vartheta = \left[(\hat{\Theta}^{(1)})^\top \cdots (\hat{\Theta}^{(r)})^\top\right]^\top$.

Consider a submatrix of (9) composed of the $L+\mu$ element of each of the r sequences of basis functions

$$W = \begin{bmatrix} \hat{\Theta}_{L+\mu-1}^{(1)} & \hat{\Theta}_{L+\mu-1}^{(2)} & \cdots & \hat{\Theta}_{L+\mu-1}^{(r)} \end{bmatrix}. \tag{10}$$

Similar to Section IV, we would like W to be invertible. Since the state at time $x_{L+\mu-1}^{(j)}$ is not a free variable, ensuring invertibility of W is a control problem. In particular, one must select the inputs such that the corresponding state and input pairs at time $L+\mu-1$ result in an invertible matrix W. In a data-driven setting, such a control problem is difficult to solve since no model knowledge is available and - to begin with - no persistently exciting data is available yet to apply data-driven control techniques. Therefore, we first show in Lemma 1 that there exist such r input sequences $u_{[0,L+\mu-1]}^{(j)}$, $j \in \mathbb{Z}_{[1,r]}$, which make W invertible and then prove in Theorem 5 how invertibility of W results in collective PE of the basis functions. Later, in the following Subsection V-A we illustrate how, under suitable assumptions, one can indeed

---

[1]This assumption is satisfied for sinusoidal functions, exponential functions and monomials, among others [23]. Note that basis functions that do not satisfy $\theta_j(\mathbf{0},\mathbf{0}) = 0$ can be suitably shifted by a constant.

find the desired control inputs a priori, which guarantee collective PE of any order for sequences of basis functions which depend on input and output data of SISO flat systems.

Lemma 1: Let Assumption 1 be satisfied and suppose that r basis functions $\theta_j : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are linearly independent on $\mathcal{R}_\mu(\mathbf{0}) \times \mathbb{R}^m$. Then there exist r sequences $u^{(j)}_{[0, L+\mu-1]}$, $j \in \mathbb{Z}_{[1,r]}$, which when applied to (8) starting from $x^{(j)}_0 = \mathbf{0}$, result in $x^{(j)}_{L+\mu-1}$, such that W in (10) is invertible.

Proof: Using similar arguments to the proof of Theorem 4, it can be shown by linear independence of the basis functions θ_j on $\mathcal{R}_\mu(\mathbf{0}) \times \mathbb{R}^m$ that there exists r pairs $(x^{(j)}_{L+\mu-1}, u^{(j)}_{L+\mu-1}) \in \mathcal{R}_\mu(\mathbf{0}) \times \mathbb{R}^m$ such that W is invertible. Since $f(\mathbf{0}, 0) = \mathbf{0}$, then starting from zero initial conditions and setting the input to $u^{(j)}_{[0, L-2]} = \mathbf{0}$, one can express the state $x^{(j)}_{L+\mu-1}$ in terms of $u^{(j)}_{[L-1, L+\mu-2]}$ only, i.e., $x^{(j)}_{L+\mu-1} = f(f \cdots (f(\mathbf{0}, u^{(j)}_{L-1}), u^{(j)}_L) \cdots, u^{(j)}_{L+\mu-2})$. Finally, since the system is locally reachable by Assumption 1, there exist inputs $u^{(j)}_{[L-1, L+\mu-1]}$, $j \in \mathbb{Z}_{[1,r]}$, which steer the system from $x^{(j)}_{L-1} = \mathbf{0}$ to $x^{(j)}_{L+\mu-1}$ in μ steps. ∎

The following theorem shows how to use the results of Lemma 1 to obtain collectively persistently exciting sequences of basis functions of any order L.

Theorem 5: Let Assumption 1 hold. Given $L, \mu \in \mathbb{Z}_{>0}$ and r basis functions $\theta_j : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ that are linearly independent on $\mathcal{R}_\mu(\mathbf{0}) \times \mathbb{R}^m$ and satisfy $\theta_j(\mathbf{0}, \mathbf{0}) = 0$, let $N_j \geq 2L + \mu - 1$ for $j \in \mathbb{Z}_{[1,r]}$. Furthermore, let the sequences $\{u^{(j)}_k\}^{N_j-1}_{k=0}$ take the form

$$u^{(j)}_k = \begin{cases} \eta^{(j)}_{[0,\mu]}, & k \in \mathbb{Z}_{[L-1, L+\mu-1]}, \\ \mathbf{0}, & \text{otherwise}, \end{cases} \qquad (11)$$

where $j \in \mathbb{Z}_{[1,r]}$, and $\eta^{(j)}_{[0,\mu]}$ are such that W in (10) is invertible. If (11) are applied to (8) starting from $x^{(j)}_0 = \mathbf{0}$, then for the matrix in (9) it holds that $\text{rank}(\mathcal{H}_L(\vartheta)) = rL$.

Proof: Each block row of (9) has r linearly independent columns given by the columns of W. Notice that each $\mathscr{H}_L(\hat{\Theta}^{(j)})$, $j \in \mathbb{Z}_{[1,r]}$, in (9) has a lower block-anti-triangular structure due to $f(\mathbf{0}, \mathbf{0}) = \mathbf{0}$, $\Theta(\mathbf{0}, \mathbf{0}) = \mathbf{0}$ and the choice of the inputs (11). As a result, every block row (9) is linearly independent from the others. Since there are L such block rows, it holds that $\text{rank}(\mathcal{H}_L(\vartheta)) = rL$. ∎

Notice that the results of Lemma 1 for locally reachable nonlinear systems are analogous to that of Theorem 4 for Hammerstein systems. However, formulating a nonlinear feasibility problem similar to (6) to find $u^{(j)}_{[0, L+\mu-1]}$, $j \in \mathbb{Z}_{[1,r]}$ would require knowledge of the unknown function f.

It was observed in simulations that randomly sampling the input sequences $\eta^{(j)}_{[0,\mu]}$, $j \in \mathbb{Z}_{[1,r]}$, in Theorem 5 from a uniform distribution typically results in a corresponding invertible matrix W. However, such a heuristic approach is not always guaranteed to achieve this result. To systematically find the desired input sequences, one must impose additional assumptions on the class of systems and the choice of basis functions. To this end, we consider in the next subsection SISO flat nonlinear systems (which are locally reachable at

the origin), and show how one can guarantee PE of any order $L > 0$ a priori, for a specific choice of basis functions.

A. SISO flat nonlinear systems

Consider an unknown SISO flat system of the form

$$x_{k+1} = f(x_k, u_k), \quad y_k = h(x_k), \qquad (12)$$

where $x_k \in \mathbb{R}^n, u_k, y_k \in \mathbb{R}$ and $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}$ are smooth unknown functions with $f(\mathbf{0}, 0) = \mathbf{0}$ and $h(\mathbf{0}) = 0$. Let $f^j_O(x_k)$ denote the j−th iterated composition of the undriven dynamics $f(x_k, 0)$.

Since the system is flat (i.e., has a well defined relative degree equal to the system dimension n, cf. [7, Sec. III.A]), it can be transformed into the discrete-time normal form provided that $0 \in \text{Im} \left(h(f^{n-1}_O(f(x, \cdot))) \right)$ holds for all $x \in \mathbb{R}^n$ (cf. [24, Sec. 2] for more details). This means that there exists an invertible (w.r.t. v_k) control law $u_k = q(x_k, v_k)$, with $q : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ and an invertible coordinate transformation $\xi_k = T(x_k) = y_{[k, k+n-1]}$, such that

$$\xi_{k+1} = A\xi_k + Bv_k, \qquad y_k = C\xi_k, \qquad (13)$$

Furthermore, A, B, C are in the Brunovsky canonical form (cf. [7, Thm. 2]) which is a controllable/observable triplet. Hence, the system is n steps locally reachable at the origin. The synthetic input v_k takes the form

$$v_k = h(f^{n-1}_O(f(x_k, u_k))). \qquad (14)$$

A sufficient condition for the analogue of Theorem 1 to flat systems [7, Prop. 1], and for designing controllers from data in [9, Cor. 2], is persistence of excitation of a sequence of basis functions which contain $h \circ f^{n-1}_O \circ f$ in their span. To check the PE condition, one typically performs an experiment of length $N \geq (r+1)L - 1$, collects the corresponding state or output measurements and then verifies the rank of the resulting Hankel matrix.

In this section, we illustrate how one can enforce PE of any order for a *particular choice* of basis functions *a priori*. A specific choice of basis functions may, in general, not contain the unknown nonlinearity (14) in its span. Nonetheless, enforcing PE of such basis functions is still useful for, e.g., designing locally stabilizing controllers for unknown SISO flat systems (cf. [9, Sec. VII.B]), and for data-driven nonlinear predictive control [13], provided that the basis functions result in a good local approximation of (14).

Since the map from u to v is invertible and since $f(\mathbf{0}, 0) = \mathbf{0}$ and $h(\mathbf{0}) = 0$, a non-zero input applied to the system from zero initial conditions results in a non-zero value of v in (14). Moreover, invertibility implies that for all $\delta_1, \delta_2 \in \mathbb{R}$, the following holds

$$\delta_1 \neq \delta_2 \iff h(f^{n-1}_O(f(\mathbf{0}, \delta_1))) \neq h(f^{n-1}_O(f(\mathbf{0}, \delta_2))). \quad (15)$$

We exploit this fact to prove the following lemma, which will be needed later for the main result of this subsection.

Lemma 2: For $t \in \mathbb{Z}_{>0}$ let $\delta_j \neq 0$, $j \in \mathbb{Z}_{[1,t]}$, be mutually distinct values and define $v_{\delta_j} := h(f^{n-1}_O(f(\mathbf{0}, \delta_j)))$. Then, the following matrix is invertible:

$$\Omega = \begin{bmatrix} v_{\delta_1} & v_{\delta_2} & \cdots & v_{\delta_t} \\ v_{\delta_1}^2 & v_{\delta_2}^2 & \cdots & v_{\delta_t}^2 \\ \vdots & \vdots & \ddots & \vdots \\ v_{\delta_1}^t & v_{\delta_2}^t & \cdots & v_{\delta_t}^t \end{bmatrix}. \tag{16}$$

Proof: Since for $j \in \mathbb{Z}_{[1,t]}$, $\delta_j \neq 0$ are mutually distinct values, it follows that the corresponding values v_{δ_j} are also distinct and non-zero (compare the discussion above Lemma 2). The matrix Ω can be written as $\Omega = V^\top \Delta$, where $V \in \mathbb{R}^{t \times t}$ is a square Vandermonde matrix composed of the distinct v_{δ_j} and, hence, invertible and $\Delta \in \mathbb{R}^{t \times t}$ is a diagonal matrix containing v_{δ_j}. The proof is concluded by noting that V and Δ are invertible matrices. ∎

In the following, we consider monomial basis functions in the transformed state and input up to some finite order $t \in \mathbb{Z}_{>0}$, and hence $r = t(n+1)$.

$$\Theta(\xi_k, u_k) = \begin{bmatrix} u_k & u_k^2 & \cdots & u_k^t & \xi_k^\top & (\xi_k^2)^\top & \cdots & (\xi_k^t)^\top \end{bmatrix}^\top. \tag{17}$$

The powers are defined element-wise, i.e., $\xi_k^t = [\xi_{1,k}^t \cdots \xi_{n,k}^t]^\top$. Notice that these functions depend only on the inputs and outputs of (12) since $\xi_k = y_{[k,k+n-1]}$ (cf. (13)). In the following theorem, we show how to choose input sequences $\{u_k^{(j)}\}_{k=0}^{N_j-1}$, $j \in \mathbb{Z}_{[1,r]}$, such that the resulting sequences of basis functions $\{\hat{\Theta}_k^{(j)}\}_{k=0}^{N_j-1}$ are collectively persistently exciting of order $L > 0$, i.e., that the corresponding mosaic Hankel matrix $\mathcal{H}_L(\vartheta)$ of the form (9) has full row rank.

Theorem 6: For $t \in \mathbb{Z}_{>0}$, let $\delta_j \neq 0$, $j \in \mathbb{Z}_{[1,t(n+1)]}$, be mutually distinct values. For $L \in \mathbb{Z}_{>0}$, $N_j \geq 2L+n-1$ and the basis functions in (17), let $\{u_k^{(j)}\}_{k=0}^{N_j-1}$ take the form in (11) with the corresponding $\eta_{[0,n]}^{(j)}$ given by

$$\eta_{[0,n]}^{(j)} = \begin{cases} \begin{bmatrix} \mathbf{0}_{n-j \times 1} \\ \delta_j \\ \mathbf{0}_{j \times 1} \end{bmatrix}, & \text{for } j \in \mathbb{Z}_{[1,n]}, \\ \vdots \\ \begin{bmatrix} \mathbf{0}_{tn-j \times 1} \\ \delta_j \\ \mathbf{0}_{j-(t-1)n \times 1} \end{bmatrix}, & \text{for } j \in \mathbb{Z}_{[(t-1)n,tn]}, \\ \begin{bmatrix} \mathbf{0}_{n \times 1} \\ \delta_j \end{bmatrix}, & \text{for } j \in \mathbb{Z}_{[tn+1,t(n+1)]}. \end{cases} \tag{18}$$

If (18) are applied to (12) starting from zero initial conditions, then $\mathrm{rank}(\mathcal{H}_L(\vartheta)) = t(n+1)L$.

Proof: Without loss of generality, let $N_j = 2L + n - 1$ for all $j \in \mathbb{Z}_{[1,t(n+1)]}$. For $c \in \mathbb{Z}_{[0,n-1]}$, we define $v_{\delta_j,[0,c]} := \begin{bmatrix} h(f_O^{n-1}(f(\mathbf{0},\delta_j))) & \cdots & h(f_O^{n+c-1}(f(\mathbf{0},\delta_j))) \end{bmatrix}^\top$ (with some abuse of notation we also use $v_{\delta_j,0} = v_{\delta_j}$). Since the system (13) is in the Brunovsky form, applying the inputs $\{u_k^{(j)}\}_{k=0}^{N_j-1}$ as defined in the theorem statement from zero initial conditions results in

$$\xi_{L+n-1}^{(j)} = \begin{cases} \begin{bmatrix} \mathbf{0}_{n-j \times 1} \\ v_{\delta_j,[0,j-1]} \end{bmatrix}, & \text{for } j \in \mathbb{Z}_{[1,n]}, \\ \vdots \\ \begin{bmatrix} \mathbf{0}_{tn-j \times 1} \\ v_{\delta_j,[0,j-(t-1)n-1]} \end{bmatrix}, & \text{for } j \in \mathbb{Z}_{[(t-1)n+1,tn]}, \\ \mathbf{0}, & \text{for } j \in \mathbb{Z}_{[tn+1,t(n+1)]}. \end{cases} \tag{19}$$

Now, we consider a submatrix of $\mathcal{H}_L(\vartheta)$ of the form of W in (10). For the choice of basis functions in (17), the inputs $\{u_k^{(j)}\}_{k=0}^{N_j-1}$ as defined in the theorem statement and the corresponding state values in (19), the matrix W takes the form (20) (see next page). Following similar arguments to the proof of Lemma 2, one can show that $W_u \in \mathbb{R}^{t \times t}$ is invertible since δ_j, $j \in \mathbb{Z}_{[tn+1,t(n+1)]}$, are non-zero and mutually distinct values. Using the columns of $W_\xi \in \mathbb{R}^{tn \times tn}$ in (20), we construct n submatrices $\overline{W}_{i,\xi} \in \mathbb{R}^{tn \times t}$, $i \in \mathbb{Z}_{[1,n]}$, of the form

$$\overline{W}_{i,\xi} =$$

$$\begin{bmatrix} \begin{pmatrix} \mathbf{0}_{n-i \times 1} \\ v_{\delta_i,[0,i-1]} \end{pmatrix} & \begin{pmatrix} \mathbf{0}_{n-i \times 1} \\ v_{\delta_{i+n},[0,i-1]} \end{pmatrix} & \cdots & \begin{pmatrix} \mathbf{0}_{n-i \times 1} \\ v_{\delta_{i+(t-1)n},[0,i-1]} \end{pmatrix} \\ \vdots & \vdots & \vdots & \vdots \\ \begin{pmatrix} \mathbf{0}_{n-i \times 1} \\ v_{\delta_i,[0,i-1]} \end{pmatrix}^t & \begin{pmatrix} \mathbf{0}_{n-i \times 1} \\ v_{\delta_{i+n},[0,i-1]} \end{pmatrix}^t & \cdots & \begin{pmatrix} \mathbf{0}_{n-i \times 1} \\ v_{\delta_{i+(t-1)n},[0,i-1]} \end{pmatrix}^t \end{bmatrix}.$$

Each matrix of this form has t rows of the form Ω in (16). Since δ_j, $j \in \mathbb{Z}_{[1,tn]}$, are non-zero and distinct, it follows from Lemma 2 that the corresponding Ω is inverible and hence, each matrix $\overline{W}_{i,\xi}$ has rank t. Notice that the columns of each matrix $\overline{W}_{i,\xi}$ are linearly independent with respect to the columns of any other $\overline{W}_{j,\xi}$, $i \neq j \in \mathbb{Z}_{[1,n]}$. This follows since (i) the rows of the form Ω appear in different rows in each $\overline{W}_{i,\xi}$ and (ii) the structure in which the block rows $\mathbf{0}_{n-i \times 1}$ appear in each submatrix. As a result, $\mathrm{rank}(W_\xi) = tn$. Due to the structure of W_u and W_ξ, it follows that $\mathrm{rank}(W) = t(n+1)$ and hence W is invertible. Finally, it follows from Theorem 5 that $\mathrm{rank}(\mathcal{H}_L(\vartheta)) = t(n+1)L$. ∎

In the following section, we illustrate the results of Theorem 6 with an example.

VI. NUMERICAL EXAMPLE

Consider a second order SISO flat system of the form in (13), with $v_k = -\sin(x_{1,k}) + x_{1,k}x_{2,k}^2 - x_{1,k}^3 x_{2,k} + u_k$. In this example, we compare the performance of three nonlinear controllers: (i) An exact linearizing and stabilizing controller designed using basis functions that include v in their span [9, Cor. 2], and two locally stabilizing controllers (ii and iii) designed using the following choice of basis functions[2] which do not contain v in their span [9, Cor. 2 and Sec. III.B]

$$\Theta(\xi_k, u_k) = \begin{bmatrix} u_k & \xi_k^\top & (\xi_k^2)^\top & (\xi_k^3)^\top \end{bmatrix}^\top. \tag{21}$$

For all three controllers, PE of the basis functions of order one is a necessary and sufficient condition for the feasibility of the convex program that is solved to obtain the control gains (cf. [9, Cor. 2, Thm. 2, and Thm. 5]). For controllers (i) and (ii), PE is enforced by sampling the input randomly. For controller (iii), PE is enforced *a priori* using the results of Theorem 6. In this case, we used a straightforward extension of [9, Cor. 2] such that collected data from multiple

---

[2]The method described in [9, Cor. 2] requires that the unknown map (14) is linear in u, which is why we use the basis functions (21). Although the choice of the basis functions in (21) is different from that in (17), one can easily see from the proof of Theorem 6 that using inputs of the form (18) also guarantees collective PE of (21).

$$W = \begin{bmatrix} \mathbf{0} & W_u \\ W_\xi & \mathbf{0} \end{bmatrix} = \left[\begin{array}{cccc|ccc} 0 & 0 & \cdots & 0 & \delta_{tn+1} & \cdots & \delta_{t(n+1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \delta^t_{tn+1} & \cdots & \delta^t_{t(n+1)} \\ \hline \begin{pmatrix} \mathbf{0}_{n-1\times 1} \\ v_{\delta_1} \end{pmatrix} & \begin{pmatrix} \mathbf{0}_{n-2\times 1} \\ v_{\delta_2,[0,1]} \end{pmatrix} & \cdots & \left(v_{\delta_{tn},[0,n-1]} \right) & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \begin{pmatrix} \mathbf{0}_{n-1\times 1} \\ v_{\delta_1} \end{pmatrix}^t & \begin{pmatrix} \mathbf{0}_{n-2\times 1} \\ v_{\delta_2,[0,1]} \end{pmatrix}^t & \cdots & \left(v_{\delta_{tn},[0,n-1]} \right)^t & \mathbf{0} & \cdots & \mathbf{0} \end{array} \right]. \tag{20}$$

TABLE I

AVERAGE LEVEL OF PE AND CUMULATIVE STABILIZATION ERROR.

Compared value	Controller (i)	Controller (ii)	Controller (iii)
avg $\sigma_{\min}(\mathcal{H}_1(\vartheta))$	0.0083	0.0083	0.0399
$\sum_{i=0}^{T-1} \frac{1}{T}|x_{1,i}|$	0.0469	0.0620	0.0553
$\sum_{i=0}^{T-1} \frac{1}{T}|x_{2,i}|$	0.0231	0.0382	0.0314

experiments (i.e., collective PE) can be used to design the controller.

Since the system is unstable, the input data (of length $N = 21$) for controllers (i) and (ii) had to be sampled from the uniform distribution $U(-0.25, 0.25)$, whereas using multiple experiments as in Theorem 6 allowed us to use inputs (each of length $N_j = 3$) with larger magnitudes (sampled from $U(-1, 1)$). In [14], a similar observation was made for linear systems. As a result, a larger quantitative level of PE was attained (cf. Remark 1 and Table I).

The performance of the closed-loop system (over $T = 20$ time instants) was compared starting from the same initial conditions (randomly sampled from $U(-1, 1) \times U(-1, 1)$). Table I shows the average cumulative stabilization errors (defined as $\sum_{k=0}^{T-1} \frac{1}{T}|x_{i,k}|$, for $i = 1, 2, T = 20$) for all three controllers over 100 experiments, excluding 5 (respectively 4) unstable experiments for controllers (ii) and (iii). Controller (i) is the best performing one since it enforces exact nonlinearity cancellation. Controller (iii) is shown to outperform controller (ii), although the same basis functions (21) were used, potentially suggesting that the region of attraction of controller (iii) is larger compared to (ii). This can be attributed to the fact that larger levels of PE were attained using multiple experiments.

VII. CONCLUSION

We provided explicit formulas for inputs that guarantee PE for linear and classes of nonlinear systems. For Hammerstein and locally reachable nonlinear systems (including SISO flat systems), we showed how to guarantee collective PE of input- and/or state-dependent basis functions. These results are crucial for the application of recent data-driven control schemes that require such PE conditions to be satisfied.

REFERENCES

[1] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice Hall, 1999.
[2] K. Åström and B. Wittenmark, *Adaptive Control*, 2nd ed. Dover Publications, 2008.
[3] M. Green and J. B. Moore, "Persistence of excitation in linear systems," *Systems & Control Letters*, vol. 7, no. 5, pp. 351–360, 1986.
[4] 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.
[5] I. Markovsky and F. Dörfler, "Behavioral systems theory in data-driven analysis, signal processing, and control," *Ann. Rev. in Control*, 2021.
[6] J. Berberich and F. Allgöwer, "A trajectory-based framework for data-driven system analysis and control," in *19th IEEE ECC*, 2020.
[7] M. Alsalti, J. Berberich, V. G. Lopez, F. Allgöwer, and M. A. Müller, "Data-based system analysis and control of flat nonlinear systems," in *60th IEEE CDC*, 2021, pp. 1484–1489.
[8] M. Alsalti, V. G. Lopez, J. Berberich, F. Allgöwer, and M. A. Müller, "Data-based control of feedback linearizable systems," *IEEE Transactions on Automatic Control*, pp. 1–8, 2023.
[9] C. De Persis, M. Rotulo, and P. Tesi, "Learning controllers from data via approximate nonlinearity cancellation," *IEEE Transactions on Automatic Control*, pp. 1–16, 2023.
[10] H. J. van Waarde, "Beyond persistent excitation: Online experiment design for data-driven modeling and control," *IEEE Control Systems Letters*, vol. 6, pp. 319–324, 2022.
[11] Z. Yuan and J. Cortés, "Data-driven optimal control of bilinear systems," *IEEE Control Systems Letters*, vol. 6, pp. 2479–2484, 2022.
[12] C. De Persis and P. Tesi, "Designing experiments for data-driven control of nonlinear systems," *IFAC-PapersOnLine*, 2021.
[13] M. Alsalti, V. G. Lopez, J. Berberich, F. Allgöwer, and M. A. Müller, "Data-driven nonlinear predictive control for feedback linearizable systems," *arXiv: 2211.06339*, accepted for 22nd IFAC WC, 2023.
[14] 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, 2020.
[15] J. Coulson, H. J. van Waarde, J. Lygeros, and F. Dörfler, "A quantitative notion of persistency of excitation and the robust fundamental lemma," *IEEE Control Systems Letters*, pp. 1–1, 2022.
[16] M. Siami, A. Olshevsky, and A. Jadbabaie, "Deterministic and randomized actuator scheduling with guaranteed performance bounds," *IEEE Transactions on Automatic Control*, vol. 66, no. 4, 2021.
[17] X. Jin, R. Bighamian, and J.-O. Hahn, "Development and in silico evaluation of a model-based closed-loop fluid resuscitation control algorithm," *IEEE Trans. Biomed. Eng.*, vol. 66, pp. 1905–1914, 2019.
[18] B. HO and R. E. Kálmán, "Effective construction of linear state-variable models from input/output functions," *at-Automatisierungstechnik*, vol. 14, no. 1-12, pp. 545–548, 1966.
[19] C. Sun and R. Dai, "Rank-constrained optimization and its applications," *Automatica*, vol. 82, pp. 128–136, 2017.
[20] I. Markovsky and K. Usevich, "Structured low-rank approximation with missing data," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 2, pp. 814–830, 2013.
[21] M. Ornik, "Guaranteed reachability for systems with unknown dynamics," in *2020 59th IEEE CDC*, 2020, pp. 2756–2761.
[22] E. D. Sontag, *Mathematical Control Theory*, 2nd ed., ser. Texts in Applied Mathematics. Springer New York, NY, 2013.
[23] O. Christensen and K. L. Christensen, "Linear independence and series expansions in function spaces," *The American Mathematical Monthly*, vol. 113, no. 7, pp. 611–627, 2006.
[24] S. Monaco and D. Normand-Cyrot, "Minimum-phase nonlinear discrete-time systems and feedback stabilization," in *26th IEEE CDC*, vol. 26, 1987, pp. 979–986.