Data-Based System Representations From Irregularly Measured Data

Mohammad Alsalti, *Graduate Student Member, IEEE*, Ivan Markovsky, *Member, IEEE*, Victor G. Lopez, *Member, IEEE*, and Matthias A. Müller

***Abstract*—Nonparametric representations of dynamical systems based on the image of a Hankel matrix of data are extensively used for data-driven control. However, if samples of data are missing, obtaining such representations becomes a difficult task. By exploiting the kernel structure of Hankel matrices of irregularly measured data generated by a linear time-invariant system, we provide computational methods for which any *complete* finite-length behavior of the system can be obtained. For the special case of periodically missing outputs, we provide conditions on the input such that the former result is guaranteed. In the presence of noise in the data, our method returns an approximate finite-length behavior of the system. We illustrate our result with several examples, including its use for approximate data completion in real-world applications and compare it to alternative methods.**

***Index Terms*—Behavioral system theory, data-based representations, irregular measurements.**

I. Introduction

MANY systems of interest are too complicated to model using first principles. However, abundance of data generated from such systems and recent advances in machine learning open new research possibilities for modeling and control design. Data- and learning-based models of dynamical systems can be obtained using, e.g., neural networks [1], Gaussian processes [2], or other machine learning techniques. Alternatively, the behavioral approach to dynamical systems theory [3] exploits system properties to arrive at nonparametric representations of dynamical systems.

A recently revived and celebrated result from the behavioral framework is the so-called fundamental lemma [4]. This lemma provides a nonparametric representation of the finite-length behavior of a discrete-time linear time-invariant (LTI) system as the image of a Hankel matrix of a single persistently exciting (PE) trajectory. In particular, it asserts that any trajectory of an LTI system is given by a linear combination of time-shifts of a single measured trajectory. This result is central to several works on direct data-based control for linear [5], [6], [7], [8] and nonlinear systems [9], [10], [11]. The reader is referred to [12] for a comprehensive review. Such direct data-based control approaches based on the image of Hankel matrices of (noisy) data have been successfully implemented on real-world systems, e.g., power systems [13], [14], heating systems [15], [16], quadcopters [17] and many others, thus, making direct data-based control an important and well-established research topic in modern control theory research.

Essential to [4] and all results that build on it is that the collected data are consecutive or made up of multiple sufficiently long trajectories of the system (see [6]). However, in practice, data can be missing due to several reasons, e.g., sensor failure and/or (irregular) packet losses in networked control systems [18]. It might also not be possible to measure some states regularly, e.g., in biomedical applications [19], [20]. This motivates the development of nonparametric representations of dynamical systems from irregularly measured data, which can then be used for system analysis and control design.

Wildhagen et al. [21] used consecutive offline input/state data of discrete-time LTI systems to design stabilizing controllers for closed-loop aperiodically sampled systems. In [22], assuming the availability of consecutive offline data, methods to complete other data sequences with missing samples are proposed by formulating data-driven approximation and interpolation problems. Both of these works still assume the availability of complete data sequences collected offline and, hence, a nonparametric model of the system is assumed to be available (cf., [4]). In contrast, the case where data are (regularly or irregularly) missing in the offline phase is not discussed. The goal of this article is to exploit the kernel structure of Hankel matrices [23], [24] in order to retrieve a nonparametric representation of the finite-length behavior of an LTI system using the available (irregularly measured) data.

In the system identification literature, missing data have been addressed in several works. The methods used include frequency domain tools [25], local optimization methods [26], and convex relaxations based on nuclear norm minimization [27]. In this work, we address the problem using the language of the

Manuscript received 24 February 2024; revised 22 May 2024; accepted 22 June 2024. Date of publication 4 July 2024; date of current version 31 December 2024. This work was supported in part by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement 948679, in part by the Catalan Institution for Research and Advanced Studies (ICREA), and in part by the Fond for Scientific Research Vlaanderen (FWO) under Project G033822 N. Recommended by Associate Editor A. Bobtsov. *(Corresponding author: Mohammad Alsalti.)*

Mohammad Alsalti, Victor G. Lopez, and Matthias A. Müller are with the Leibniz University Hannover, Institute of Automatic Control, 30167 Hannover, Germany (e-mail: alsalti@irt.uni-hannover.de; lopez@irt.uni-hannover.de; mueller@irt.uni-hannover.de).

Ivan Markovsky is with the International Centre for Numerical Methods in Engineering, 08860 Barcelona, Spain, and also with the Catalan Institution for Research and Advanced Studies, 08860 Barcelona, Spain (e-mail: imarkovsky@cimne.upc.edu).

Digital Object Identifier 10.1109/TAC.2024.3423053

1558-2523 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

behavioral framework [3], as was done in [28] and [29]. There, a dynamical system is viewed as a set of trajectories, allowing us to define the system's properties in terms of its behavior. A representation of the system, however, is defined independently from its behavior and one of several representations of a dynamical system is the kernel representation. An algorithm for identifying a *minimal* kernel representation of scalar systems despite missing data was proposed in [28] and [29]. However, no conditions were given to guarantee that this algorithm succeeds for any pattern of missing data.

The contributions of this article are as follows.

1) We first modify the algorithm in [28] and [29] and use it to obtain a kernel representation of multivariable systems using irregularly measured data.
2) We then exploit the kernel structure of Hankel matrices to recover any *complete* finite-length behavior of the system, i.e., we recover a nonparametric model of the system similar to the ones given by the fundamental lemma [4] and its generalization [30]. Different from existing works, this is done using only irregularly measured offline data.
3) We consider the special case of periodically missing outputs and provide sufficient conditions on the input such that a kernel representation is guaranteed to be retrieved from the available data. This allows us to propose a fundamental-lemma-like result using data with periodically missing outputs.
4) We investigate the presence of noise in the data and provide computational methods to obtain an approximate finite-length behavior of the system.

The rest of this article is organized as follows. Section II introduces the notation and reviews the preliminary results. Section III formulates the problem of finding kernel representations from missing data, and includes the corresponding algorithm required to compute them. Section IV shows how to use these representations in order to obtain a nonparametric representation of any *complete* finite-length behavior of the system, by exploiting the kernel structure of Hankel matrices. Section V focuses on the special case of periodically missing outputs and provides conditions on the input such that a fundamental-lemma-like result is obtained. In Section VI we propose a modification of our results to handle the case of noisy data, such that an approximate finite-length behavior of the system is obtained. Section VII verifies the results using numerical examples and illustrates how the resulting representation can be used for approximate data completion in real-world applications, including comparisons to alternative methods. Finally, Section VIII concludes this article.

II. Preliminaries

The sets of integers, natural and real numbers are denoted by $\mathbb{Z}, \mathbb{N}, \mathbb{R}$, respectively. The restriction of integers to an interval is denoted by $\mathbb{Z}_{[a,b]}$ for $b > a \in \mathbb{Z}$. We use I_m to denote an $m \times m$ identity matrix and $0_{n\times m}$ to denote an $n \times m$ matrix of zeros; when the dimensions are clear from the context, we omit the subscript for simplicity. We use $\lfloor z \rfloor$ to denote the floor operator which acts on $z \in \mathbb{R}$ and returns the greatest integer less than or equal to z. For a matrix $M \in \mathbb{R}^{m\times n}$, we denote its image by $\text{im}(M)$ and its kernel by $\ker(M)$. When a basis of $\ker(M)$ is to be computed, we write $N = \text{null}(M)$, which returns a matrix N of appropriate dimensions such that $MN = 0$. The set of polynomial matrices (in z) with coefficients in $\mathbb{R}^{m\times n}$ is denoted by $\mathbb{R}^{m\times n}[z]$. For a subspace $\mathcal{A}$ of $\mathbb{R}^n$, we denote its orthogonal complement by $\mathcal{A}^\perp$. Further, for another subspace $\mathcal{B}$ of $\mathbb{R}^n$ such that $\mathcal{A} \cap \mathcal{B} = \{0\}$, the direct sum of the two subspaces is denoted by $\mathcal{A} \oplus \mathcal{B}$.

The set of infinite length $q-$variate, real-valued time series $w = (w(0), w(1), \ldots)$ is denoted by $(\mathbb{R}^q)^{\mathbb{N}}$. For $T \in \mathbb{N}$, the set of finite-length $q-$variate, real-valued time series $w = (w(0), w(1), \ldots, w(T-1))$ is denoted by $(\mathbb{R}^q)^T$. With slight abuse of notation, we also use w to denote the stacked vector of the time series as

$$w = \begin{bmatrix} (w(0))^\top & (w(1))^\top & \cdots & (w(T-1))^\top \end{bmatrix}^\top \in \mathbb{R}^{qT}$$

and its restriction to L as

$$w_{[0,L-1]} = \begin{bmatrix} (w(0))^\top & (w(1))^\top & \cdots & (w(L-1))^\top \end{bmatrix}^\top \in \mathbb{R}^{qL}.$$

The Hankel matrix of depth L of w is defined as

$$\mathscr{H}_L(w) = \begin{bmatrix} w_{[0,L-1]} & w_{[1,L]} & \cdots & w_{[T-L,T-1]} \end{bmatrix}.$$

The behavior $\mathscr{B}$ of a dynamical system is defined as a set of infinite-length trajectories (cf., [3]). The finite-length behavior of a system is a set of finite-length trajectories. For $T \in \mathbb{N}$, this is denoted by $\mathscr{B}|_T$ and, hence, a trajectory of length T of the system is denoted by $w \in \mathscr{B}|_T \subset (\mathbb{R}^q)^T$. The system is linear if $\mathscr{B}$ is a subspace (of $(\mathbb{R}^q)^{\mathbb{N}}$) and it is time-invariant if it is invariant to the action of the *shift operator*, defined as $\sigma : (\mathbb{R}^q)^{\mathbb{N}} \to (\mathbb{R}^q)^{\mathbb{N}}$ such that $\sigma^j w(k) := w(k+j)$ for $j \in \mathbb{N}$.

Let $u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}^p$ denote the inputs and outputs of system $\mathscr{B}$ at time t. For a permutation matrix $\Pi \in \mathbb{R}^{q\times q}$, we define a partitioning of the variable $w(t) \in \mathbb{R}^q$ such that $w(t) = \Pi \begin{bmatrix} u(t) \\ y(t) \end{bmatrix}$. The set of discrete-time LTI systems with $q-$variables and fixed complexity (m, n, ℓ) is denoted by $\partial\mathscr{L}^q_{m,n,\ell}$, where $q = m + p$ and (m, n, ℓ) denote: 1) the number of inputs $\mathbf{m}(\mathscr{B}) = m$; 2) the order of the system $\mathbf{n}(\mathscr{B}) = n$; 3) the lag of the system $\boldsymbol{\ell}(\mathscr{B}) = \ell$, which is the observability index in the state-space framework. These integers satisfy the following relation $\ell \leq n \leq p\ell$ [30].

A finite-dimensional LTI system $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ admits a kernel representation [3, Part I]

$$\mathscr{B} = \ker R(\sigma) := \{w : \mathbb{N} \to \mathbb{R}^q \mid R(\sigma)w = 0\} \tag{1}$$

where the operator $R(\sigma)$ is defined by the polynomial matrix

$$\begin{aligned} R(z) &= R_0 + R_1 z + \cdots + R_\ell z^\ell \\ &= \begin{bmatrix} r_1(z) \\ \vdots \\ r_g(z) \end{bmatrix} = \begin{bmatrix} r_{1,0} + r_{1,1}z + \ldots + r_{1,\ell_1}z^{\ell_1} \\ \vdots \\ r_{g,0} + r_{g,1}z + \ldots + r_{g,\ell_g}z^{\ell_g} \end{bmatrix} \in \mathbb{R}^{g\times q}[z] \end{aligned} \tag{2}$$

with $r_{i,j} \in \mathbb{R}^{1\times q}$. A *minimal* kernel representation is one where the degree of the polynomial matrix R is minimal. Each row

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

$r_i(z)$ of $R(z)$ defines an *annihilator* $r(\sigma)$ of the system and is formally defined as follows.

Definition 1: [30, Def. 2] An annihilator of $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ is an operator $r(\sigma)$, $r(z) \in \mathbb{R}^{1\times q}[z]$, such that $r(\sigma)\mathscr{B} = 0$.

A finite dimensional LTI system $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ also admits an input/state/output representation of the form

$$\mathscr{B} = \mathscr{B}_{ss}(A,B,C,D,\Pi) := \left\{ \Pi \begin{bmatrix} u \\ y \end{bmatrix} \mid \exists x \in (\mathbb{R}^n)^{\mathbb{N}} \right.$$
$$\left. \text{s.t.} \quad \sigma x = Ax + Bu,\ y = Cx + Du \right\}. \tag{3}$$

One of the key ideas in the behavioral framework is the notion of the *most powerful unfalsified model* (MPUM) of the data [3, Part II], denoted $\mathscr{B}_{\text{MPUM}}$. The MPUM always exists and is unique. Its restricted behavior, i.e., $\mathscr{B}_{\text{MPUM}}|_L$, is given by the image of the Hankel matrix constructed from data. For instance, let $w \in \mathscr{B}|_T$, where $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$, then $\mathscr{B}_{\text{MPUM}}|_L = \text{im}(\mathscr{H}_L(w))$. Furthermore, due to linearity and shift-invariance, the following holds $\mathscr{B}_{\text{MPUM}}|_L = \text{im}(\mathscr{H}_L(w)) \subseteq \mathscr{B}|_L$, where $\mathscr{B}|_L$ is the restricted behavior of the system. This is summarized in the following theorem, which further gives necessary and sufficient conditions under which the equality holds, i.e., $\text{im}(\mathscr{H}_L(w)) = \mathscr{B}|_L$. Theorem 1 follows from the generalized persistency of excitation result [30, Cor. 21], which generalizes the fundamental lemma [4, Th. 1] to uncontrollable systems and does not require a priori known input–output partitioning of the data, nor persistency of excitation of the input (see also [31]).

Theorem 1: [30] Let $w \in \mathscr{B}|_T$ with $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$. Then, for all $L \in \mathbb{Z}_{[\ell, L_{\max}]}$ where $L_{\max} := \lfloor \frac{T+1}{q+1} \rfloor$, the following holds:

$$\text{im}\left(\mathscr{H}_L(w)\right) \subseteq \mathscr{B}|_L. \tag{4}$$

Moreover, for $L \geq \ell$, $\text{im}(\mathscr{H}_L(w)) = \mathscr{B}|_L$ if and only if

$$\text{rank}(\mathscr{H}_L(w)) = mL + n. \tag{GPE}$$

When (GPE) holds, one obtains a *nonparametric data-based representation* of the restricted behavior of the system $\mathscr{B}|_L$ for $L \geq \ell$, i.e., any trajectory of the system is given by a linear combination of the columns of the Hankel matrix $\mathscr{H}_L(w)$. This is formalized in the following corollary.

Corollary 1: [22, Cor. 4] Let $w \in \mathscr{B}|_T$ with $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ and let (GPE) hold for $L \geq \ell$. Then $\bar{w} \in \mathscr{B}|_L$ if and only if there exists $\alpha \in \mathbb{R}^{T-L+1}$ such that

$$\mathscr{H}_L(w)\alpha = \bar{w}. \tag{5}$$

Another result which follows from the rank condition in (GPE) is the following. It allows us to retrieve a kernel representation (2) of an LTI system directly from data. This is formalized in the following corollary.

Corollary 2: Let $w \in \mathscr{B}|_T$ where $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$. Let $d \geq \ell + 1$, and suppose

$$\text{rank}(\mathscr{H}_d(w)) = md + n. \tag{6}$$

Then, the coefficients of the polynomial matrix R in (2) are given by the basis of the left null space of the Hankel matrix $\mathscr{H}_d(w)$, i.e., R is defined in terms of the rows of $R_d \in \mathbb{R}^{pd-n\times qd}$ where

$$R_d = \begin{bmatrix} r_{1,0} & r_{1,1} & \cdots & r_{1,d-1} \\ r_{2,0} & r_{2,1} & \cdots & r_{2,d-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_{pd-n,0} & r_{pd-n,1} & \cdots & r_{pd-n,d-1} \end{bmatrix} \tag{7}$$

with $r_{i,j} \in \mathbb{R}^{1\times q}$ which satisfies $R_d\mathscr{H}_d(w) = 0$.

Proof: Existence of R_d (of proper dimension) follows from (6). Since (6) holds, then $\text{im}(\mathscr{H}_d(w)) = \mathscr{B}|_d$ (by Theorem 1). Therefore, the rows of the matrix R_d represent the coefficients of the annihilators of $\mathscr{B}|_d$ (see Definition 1). Since the map $\mathscr{B}|_d \mapsto \mathscr{B}$ is well defined (see [30, Lemma 13]), then the rows of R_d also represent the coefficients of the annihilators of $\mathscr{B}$ and, hence, define a kernel representation as in (2) with $g = pd - n$ and $\ell_i = d - 1$ for all $i = \{1, \ldots, pd - n\}$. □

One of the goals of this article is to arrive at a result similar to that of Corollary 1 when the data is not consecutive (i.e., irregularly measured or contains missing values). To this end, we denote a missing data value by NaN (short for not a number). This object is defined such that $0 \cdot \text{NaN} = 0$ and $1 \cdot \text{NaN} = \text{NaN}$. Furthermore, we define an extended set of the real numbers as the union of the set of $\mathbb{R}$ and NaN as $\mathbb{R}_{\text{ext}} := \mathbb{R} \cup \text{NaN}$. Throughout the article, we denote a trajectory of $\mathscr{B}|_T$ by $w \in (\mathbb{R}^q)^T$, and its corresponding *irregular measurements* by $w^- \in (\mathbb{R}^q_{\text{ext}})^T$, with NaNs appearing arbitrarily in w^-, e.g., $w^-(t) = \begin{bmatrix} w_1^-(t) & \cdots & \text{NaN} & \cdots & w_q^-(t) \end{bmatrix}^\top$, for some $t \in \mathbb{Z}_{[0,T-1]}$. This means that we allow for portions of a sample or entire samples of data to be missing. For the majority of the subsequent developments, we consider *exact* data, i.e., noise-free data. Later in Section VI, we consider the case of noisy data.

III. Kernel Representations From Missing Data

A. Problem Formulation

Whether in Theorem 1, Corollary 1 or Corollary 2, if the data w are not complete (containing missing values), then one cannot evaluate (GPE) or (6), nor use (5) as a nonparametric model anymore. One can instead use multiple (short) experiments to build a *mosaic Hankel matrix* (cf., [28]). In particular, given $\mathscr{W} = \{w^{(1)}, w^{(2)}, \ldots, w^{(N)}\}$ with each $w^{(i)} \in \mathscr{B}|_{T_i}$, for $N, T_i \in \mathbb{N}$, $i \in \mathbb{Z}_{[1,N]}$, one can construct

$$\mathcal{H}_L(\mathscr{W}) := \begin{bmatrix} \mathscr{H}_L\left(w^{(1)}\right) & \cdots & \mathscr{H}_L\left(w^{(N)}\right) \end{bmatrix}$$

for $\ell \leq L \leq T_{\min} := \min_i T_i$. Later, it was shown in [30] that if $\text{rank}(\mathcal{H}_L(\mathscr{W})) = mL + n$, then any $\bar{w} \in \mathscr{B}|_L$ if and only if there exists $\alpha \in \mathbb{R}^{\sum_{i=1}^N (T_i - L + 1)}$ such that $\mathcal{H}_L(\mathscr{W})\alpha = \bar{w}$. Similar ideas were presented in [6], but only controllable systems were addressed, an a priori known input–output partitioning was required and a collective persistency of excitation condition on the input sequences $u^{(i)}$ was imposed. However, both [6], [30] require that each trajectory $w^{(i)}$ is composed of at least L consecutive samples.

In this work, we provide computational methods to retrieve a nonparametric model as in (5) from data which has (regularly

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

or irregularly) missing values. We do not require that data is consecutive for L samples, nor do we impose controllability or persistency of excitation. To do so, we modify the algorithm in [28] and [29] to compute a (nonminimal) kernel representation of the system. For convenience, we first illustrate the procedure with the following (simple) motivating example and later formalize the algorithm in Section III-B. Note that none of the following steps have to be carried out by the user, instead all steps are implemented in a single algorithm (see Algorithm 1).

Example 1: Consider the system $w(t) = 2w(t-1) - w(t-2)$. Here, $n = 2$, $\ell = 2$, $p = 1$, $m = 0$. Suppose an experiment of length $T = 8$ was performed on the system from initial conditions $w(0) = 1$ and $w(1) = 2$. The following are the complete and missing data sequences, respectively:

$$w = \{1, 2, 3, 4, 5, 6, 7, 8\}$$
$$w^- = \{1, 2, \text{NaN}, 4, 5, \text{NaN}, 7, 8\}.$$

The data are missing periodically with period equal to $\ell + 1$. This prevents us from obtaining the kernel representation as in Corollary 2 since

$$\mathscr{H}_{\ell+1}(w^-) = \begin{bmatrix} 1 & 2 & \text{NaN} & 4 & 5 & \text{NaN} \\ 2 & \text{NaN} & 4 & 5 & \text{NaN} & 7 \\ \text{NaN} & 4 & 5 & \text{NaN} & 7 & 8 \end{bmatrix}$$

has missing values in every row and every column. Instead, one constructs an extended series by concatenating w with $T-$long instances of NaN, i.e.,

$$w_{\text{ext}} = (\underbrace{1, 2, \text{NaN}, 4, 5, \text{NaN}, 7, 8}_{w^-}, \underbrace{\text{NaN}, \ldots, \text{NaN}}_{T \text{ times}})$$

where $T = 8$. This extended vector allows us to investigate the following Hankel matrix, which encompasses all other shallower Hankel matrices of w^-:

$$\mathscr{H}_T(w_{\text{ext}}) = \left[\begin{array}{cccccccc} 1 & 2 & \text{NaN} & 4 & 5 & \text{NaN} & 7 & 8 \\ 2 & \text{NaN} & 4 & 5 & \text{NaN} & 7 & 8 & \text{NaN} \\ \text{NaN} & 4 & 5 & \text{NaN} & 7 & 8 & \text{NaN} & \text{NaN} \\ 4 & 5 & \text{NaN} & 7 & 8 & \text{NaN} & \text{NaN} & \text{NaN} \\ \hline 5 & \text{NaN} & 7 & 8 & \text{NaN} & \text{NaN} & \text{NaN} & \text{NaN} \\ \text{NaN} & 7 & 8 & \text{NaN} & \text{NaN} & \text{NaN} & \text{NaN} & \text{NaN} \\ 7 & 8 & \text{NaN} & \text{NaN} & \text{NaN} & \text{NaN} & \text{NaN} & \text{NaN} \\ 8 & \text{NaN} & \text{NaN} & \text{NaN} & \text{NaN} & \text{NaN} & \text{NaN} & \text{NaN} \end{array}\right]. \tag{8}$$

As in Corollary 2, when $\text{rank}(\mathscr{H}_d(w)) = md + n$, the kernel representation of the system is given by the rows of R_d where $R_d\mathscr{H}_d(w) = 0$ for some $d \geq \ell + 1$. The key step is to select submatrices of $\mathscr{H}_d(w^-)$ that do not have NaN values but have nontrivial left kernels. For instance, consider

$$H_1 = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ \text{NaN} & \text{NaN} \\ 4 & 7 \end{bmatrix}, \; H_2 = \begin{bmatrix} 2 & 5 \\ \text{NaN} & \text{NaN} \\ 4 & 7 \\ 5 & 8 \end{bmatrix}, \; H_3 = \begin{bmatrix} \text{NaN} \\ 4 \\ 5 \\ \text{NaN} \end{bmatrix}$$

which are found by taking the first and fourth, second and fifth, and third columns of $\mathscr{H}_{\ell+2}(w^-)$, respectively (here, $d = \ell + 2$ and $\mathscr{H}_{\ell+2}(w^-)$ is the indicated top left submatrix of (8)). By deleting the shared NaN rows of H_1, H_2, H_3, one can find bases for the nontrivial left kernels of the resulting matrices $\bar{H}_1, \bar{H}_2, \bar{H}_3$ as

$$\underbrace{\begin{bmatrix} 1 & -3/2 & 1/2 \end{bmatrix}}_{\bar{Z}_1} \underbrace{\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 4 & 7 \end{bmatrix}}_{\bar{H}_1} = 0, \quad \underbrace{\begin{bmatrix} 1 & -3 & 2 \end{bmatrix}}_{\bar{Z}_2} \underbrace{\begin{bmatrix} 2 & 5 \\ 4 & 7 \\ 5 & 8 \end{bmatrix}}_{\bar{H}_2} = 0$$

$$\underbrace{\begin{bmatrix} -5/4 & 1 \end{bmatrix}}_{\bar{Z}_3} \underbrace{\begin{bmatrix} 4 \\ 5 \end{bmatrix}}_{\bar{H}_3} = 0.$$

When extending $\bar{Z}_i$ to Z_i by inserting zeros in place of the deleted NaN rows, the following holds by definition of NaN:

$$\underbrace{\begin{bmatrix} 1 & -3/2 & 0 & 1/2 \end{bmatrix}}_{Z_1} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ \text{NaN} & \text{NaN} \\ 4 & 7 \end{bmatrix} = 0$$

$$\underbrace{\begin{bmatrix} 1 & 0 & -3 & 2 \end{bmatrix}}_{Z_2} \begin{bmatrix} 2 & 5 \\ \text{NaN} & \text{NaN} \\ 4 & 7 \\ 5 & 8 \end{bmatrix} = 0,$$

$$\underbrace{\begin{bmatrix} 0 & -5/4 & 1 & 0 \end{bmatrix}}_{Z_3} \begin{bmatrix} \text{NaN} \\ 4 \\ 5 \\ \text{NaN} \end{bmatrix} = 0.$$

Consider a vector $\bar{Z}_i$ in the left kernel of a submatrix $\bar{H}_i$. In order for the extended vector Z_i to be in the left kernel of $\mathscr{H}_{\ell+2}(w)$, we require that $\bar{H}_i$ and $\mathscr{H}_{\ell+2}(w)$ have the same rank (see Lemma 1 below). In particular

$$\text{rank}(\bar{H}_i) = \text{rank}(\mathscr{H}_{\ell+2}(w)) \Rightarrow \left(\bar{Z}_i^\top \in \ker(\bar{H}_i^\top) \Rightarrow Z_i^\top \in \ker(\mathscr{H}_{\ell+2}^\top(w))\right). \tag{9}$$

We point out that $\text{rank}(\mathscr{H}_{\ell+2}(w))$ cannot be evaluated since w is not completely known. However, for existence of $R_{\ell+2}$ we require that $\text{rank}(\mathscr{H}_{\ell+2}(w)) = m(\ell+2) + n$ (compare Corollary 2). Since the complexity is known, one can test whether $\text{rank}(\bar{H}_i) = m(\ell+2) + n$ as the matrices $\bar{H}_i$ do not have missing values.

For this example, $\text{rank}(\bar{H}_1) = \text{rank}(\bar{H}_2) = m(\ell+2) + n = 2$, whereas $\text{rank}(\bar{H}_3) = 1$. From (9), it holds that $\mathscr{Z}\mathscr{H}_{\ell+2}(w) = 0$, where

$$\mathscr{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} 1 & -3/2 & 0 & 1/2 \\ 1 & 0 & -3 & 2 \end{bmatrix}. \tag{10}$$

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

Finally, a basis for the kernel representation $R_{\ell+2}$ is given by the $p(\ell+2)-n=2$ linearly independent rows in $\mathscr{Z}$. For this example, $\mathscr{Z}=R_{\ell+2}$.

The steps followed in Example 1 are a modification of the results of [28]. Unlike [28], our method finds a (not necessarily minimal) kernel representation and works for multivariable systems. The most significant modification was to impose additional rank constraints on the selected submatrices (see (9) and Lemma 1 below) such that the output of the algorithm is indeed a kernel representation of the system. This is an important condition for the selected submatrices to be *informative* in the sense that the left kernel of each submatrix reveals a part of the kernel representation of $\mathscr{B}$ (compare (GPE) and Corollary 2). In the next section, we formalize the steps of Example 1. Later in Section V, we consider a special case of periodically missing outputs and provide conditions on the input such that the algorithm is guaranteed, a priori, to return the kernel representation.

B. Algorithm for Computing the Kernel Representation

In this section, we formalize the steps followed in Example 1. We do not make any assumption on the pattern of missing data, and only assume knowledge of the model complexity. We start by stating the following lemma, which formalizes the rank condition (9). This lemma is needed in our modification of the algorithm in [28] (see Algorithm 1 below). Specifically, it represents a condition on the selected submatrix to be informative in the sense that its left kernel reveals a part of the system's annihilators.

Lemma 1: Let $A\in\mathbb{R}^{n\times m}$ be a block matrix of the form

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}$$

with $A_{11}\in\mathbb{R}^{n_1\times m_1}$, $A_{12}\in\mathbb{R}^{n_1\times m_2}$, $A_{21}\in\mathbb{R}^{n_2\times m_1}$, $A_{22}\in\mathbb{R}^{n_2\times m_2}$ and where $n=n_1+n_2$ and $m=m_1+m_2$. Suppose A has a nontrivial left kernel, i.e., $\operatorname{rank}(A)<n$. If $\operatorname{rank}(A_{11})=\operatorname{rank}(A)$ and A_{11} has a nontrivial left kernel, then

$$\bar{z}A_{11}=0\Rightarrow zA=0\qquad \forall \bar{z}^\top\in\ker\left(A_{11}^\top\right)\tag{11}$$

where $z=[\bar{z}\quad 0_{1\times n_2}]\in\mathbb{R}^{1\times n}$.

Proof: Since $\operatorname{rank}(A_{11})=\operatorname{rank}(A)$ and A_{11} has a nontrivial left kernel, there exists $\bar{z}^\top\in\mathbb{R}^{n_1}$ such that $\bar{z}A_{11}=0$. As in the lemma statement, let $z=[\bar{z}\quad 0_{1\times n_2}]$. By construction, it holds that

$$z\begin{bmatrix}A_{11}\\ A_{21}\end{bmatrix}=[\bar{z}\quad 0_{1\times n_2}]\begin{bmatrix}A_{11}\\ A_{21}\end{bmatrix}=\bar{z}A_{11}+0\cdot A_{21}=0.\tag{12}$$

Furthermore, since $\operatorname{rank}(A_{11})=\operatorname{rank}(A)$, it holds that the columns of A_{12} can be written as a linear combination of the columns of A_{11}. This means that there exists a matrix $K\in\mathbb{R}^{m_1\times m_2}$ such that $A_{12}=A_{11}K$ and, hence

$$z\begin{bmatrix}A_{12}\\ A_{22}\end{bmatrix}=[\bar{z}\quad 0_{1\times n_2}]\begin{bmatrix}A_{11}K\\ A_{22}\end{bmatrix}=\bar{z}A_{11}K+0\cdot A_{22}=0.\tag{13}$$

Algorithm 1: Computing a Kernel Representation.

Input: Irregular measurements $w^-\in(\mathbb{R}^q_{\text{ext}})^T$ of a trajectory $w\in\mathscr{B}|_T$ where $\mathscr{B}\in\partial\mathscr{L}^q_{m,n,\ell}$ is of known complexity.

1. Construct the extended sequence w_{ext} by appending w^- with T instances of NaN.
2. Construct the matrix $\mathscr{H}_T(w_{\text{ext}})$, and initialize $\mathscr{Z}=[\,]$.
3. **For** $\ell+1\le d\le T$ **do**
   - Search over all submatrices of $\mathscr{H}_d(w^-)$, i.e., $\bar{H}_i\in\mathbb{R}^{\eta_i\times\rho_i}$, that have $\operatorname{rank}(\bar{H}_i)=md+n$ and nontrivial left kernels i.e., $\eta_i>md+n$.
   - Compute a basis for $\ker(\bar{H}_i^\top)$, i.e., a full row rank matrix $\bar{Z}_i\in\mathbb{R}^{\zeta_i\times\eta_i}$ such that $\bar{Z}_i\bar{H}_i=0$.
   - Extend $\bar{Z}_i$ to Z_i by inserting zero columns at the location of the rows removed from $\mathscr{H}_d(w^-)$ to obtain $\bar{H}_i$.
   - Concatenate all the above matrices Z_i in $\mathscr{Z}$, i.e., $\mathscr{Z}:=\begin{bmatrix}[\mathscr{Z}\quad \mathbf{0}]^\top & Z_1^\top & Z_2^\top & \cdots\end{bmatrix}^\top$, where $\mathbf{0}$ denotes a matrix of zeros of appropriate dimensions.
   - **If** $\operatorname{rank}(\mathscr{Z})=pd-n$ **break**.

   end for

Output: The matrix R_d specifying the kernel representation is given by any $pd-n$ linearly independent rows of $\mathscr{Z}$.

Equations (12) and (13) together imply that $zA=0$, which completes the proof. □

Remark 1: The rows and columns of A_{ij} in Lemma 1 need not appear consecutively in the matrix A. This is because row and column permutations do not change the rank of A. To see how Lemma 1 generalizes (9) in Example 1, one can carry out the necessary row/column permutations that bring the submatrix $\bar{H}_i$ in $\mathscr{H}_{\ell+2}(w^-)$ to the top left corner, i.e., the location of A_{11} in the matrix A in Lemma 1.

Algorithm 1 generalizes the steps followed in Example 1. The algorithm takes as an input a sequence of irregularly measured data w^- and does not require any further steps to be performed by the user. Then, it searches over different depths of $\mathscr{H}_T(w_{\text{ext}})$ for possible submatrices $\bar{H}_i$ that have nontrivial left kernels (checkable by a rank condition). According to Lemma 1 (compare also (9) and Remark 1), suitable extensions of vectors in the left null space of the aforementioned $\bar{H}_i$ are guaranteed to form annihilators of the system. These annihilators are collected in the matrix[1] $\mathscr{Z}$. If enough annihilators are found, or if the maximum depth of the Hankel matrix is reached, the algorithm terminates. The algorithm is deemed successful if $\mathscr{Z}$ contains $pd-n$ linearly independent rows (where $pd-n$ is the dimension of the left kernel of $\mathscr{H}_d(w)$). These rows specify a kernel representation of the system (cf., Corollary 2). If $\operatorname{rank}(\mathscr{Z})<pd-n$, then the algorithm did not succeed in returning the kernel representation.

[1]Notice that the matrix $\mathscr{Z}$ is adapted at each iteration such that information from previous iterations is taken into account.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

At this point, we do not impose any conditions on the pattern of the missing data. As a result, we cannot a priori guarantee whether or not this algorithm will succeed or fail. The algorithm may fail, for instance, if there are not enough submatrices $\bar{H}_i$ that have the desired rank (i.e., data are not informative). However, even if the algorithm fails, it still reveals a part of the true system representation given by the annihilators which the algorithm manages to find. In general, the success rate of Algorithm 1 increases as the length of the sequence T increases and as the fraction of given to missing samples increases as well. This is illustrated by an example in Section VII-A.

In some cases, prior knowledge on the pattern of missing data is available and one can exploit it to guarantee successful completion of Algorithm 1. In Section V, we consider the case of periodically missing output samples and show sufficient conditions (including a PE condition on the input and a lower bound for the length of the sequence T) such that Algorithm 1 is guaranteed to return a kernel representation of the system.

We point out that the most expensive step of Algorithm 1 is the rank check and computation of a basis for the left kernel of submatrices $\bar{H}_i$. This corresponds to a singular value decomposition of that matrix, which has complexity $O(\max\{\eta_i,\rho_i\}(\min\{\eta_i,\rho_i\})^2)$ (cf., [32]), where η_i and ρ_i are the dimensions of the complete submatrix $\bar{H}_i$ of the Hankel matrix $\mathscr{H}_d(w^-)$. Later in Section VII, we illustrate with numerical examples that Algorithm 1 takes less time to complete compared to other methods that rely on data completion (which requires solving nonlinear or convex programs, e.g., convex relaxations based on nuclear norm minimization [27]) followed by finding a basis for the left kernel.

In the following section, we use the output of Algorithm 1 to obtain the *complete* finite-length behavior of the system $\mathscr{B}|_L$, for *any* $L \geq d$.

IV. Retrieving a Nonparametric System Representation

Once a kernel representation of an LTI system is obtained, e.g., by Algorithm 1, we use this kernel representation to reconstruct the restricted behavior of the system. This can then be used as a nonparametric model of the system similar to that in Corollary 1. We do this by exploiting the kernel structure of Hankel matrices [23], [24].

A. Kernel Structure of Hankel Matrices

We start by reviewing the concept of *shift-chains*.

Definition 2 ([23]): A shift-chain of length r generated by a vector $z \in \mathbb{R}^s$ with $z \neq 0$ is a set of vectors $v_0, \ldots, v_{r-1} \in \mathbb{R}^n$, with $s + r = n + 1$, such that $v_k = \Pi_{n,s}^k z \quad \forall k \in \mathbb{Z}_{[0,r-1]}$ where $\Pi_{n,s}^k$ is a permutation matrix having n rows, s columns and k specifying the nonzero diagonal.

To better illustrate the definition of a shift-chain, consider the following example.

Example 2: Let $z = [a \quad b \quad c]^\top \in \mathbb{R}^3$ and let $n = 5$ (hence, $r = n + 1 - s = 3$). The following vectors form a shift-chain:

$$v_0 = [a \quad b \quad c \quad 0 \quad 0]^\top, \quad v_1 = [0 \quad a \quad b \quad c \quad 0]^\top$$

$$v_2 = [0 \quad 0 \quad a \quad b \quad c]^\top.$$

The authors in [23] and [24] investigated the kernel structure of Hankel matrices. Being a structured matrix, its kernels posses certain structures as well. The following theorem shows how the kernels of a Hankel matrix can be represented using shift-chains.

Theorem 2: [24, Cor. 2.3] For a sequence $t = (t_0, t_1, \ldots)$, where $t_i \in \mathbb{R}^{p\times q}$, the (left or right) kernel of a Hankel matrix of t is given by the linear hull of at most $p + q$ shift-chains.

In the next section, we construct a shift-chain-like matrix out of the computed kernel representation obtained from Algorithm 1 and use it to later obtain the *complete* restricted behavior (of *any* length $L \geq d \geq \ell + 1$) of the LTI system. This is done by showing that the shift-chain-like matrix forms a basis for the left kernel of any deeper Hankel matrix (whose image is equal to $\mathscr{B}|_L$) and then using orthogonality of the associated subspaces of the Hankel matrix to retrieve the restricted behavior of the system.

B. Retrieving the Restricted Behavior of an LTI System

In Section III-B, we used Algorithm 1 to obtain a kernel representation of the system using irregularly measured data. If successful, the algorithm returns a full row rank matrix $R_d \in \mathbb{R}^{pd-n\times qd}$ such that $R_d\mathscr{H}_d(w) = 0$. In contrast, Theorem 1 provides a nonparametric model of the restricted behavior of an LTI system as the image of a Hankel matrix consisting of the data generated by the system $\mathscr{H}_L(w)$, provided that this matrix satisfies the rank condition (GPE) (compare [4] and [30]). In this section, we use a kernel representation R_d to find a complete matrix (i.e., without missing values) that has the same rank and image of the matrix $\mathscr{H}_L(w)$, thus corresponding to a nonparametric model of the restricted behavior of the true data-generating system. Furthermore, we show how one can retrieve the *complete* restricted behavior of the system $\mathscr{B}|_L$ for *any* $L \geq d \geq \ell + 1$, i.e., *independent of the number of available data points*. We start with an illustrative example using the same system as in Example 1.

Example 3: Consider again the system and setting of Example 1. There, it was shown that one can use w^- to obtain a basis for the left kernel of $\mathscr{H}_{\ell+2}(w)$, i.e., a matrix $R_{\ell+2}$ which satisfies $R_{\ell+2}\mathscr{H}_{\ell+2}(w) = 0$ and takes the form

$$R_{\ell+2} = \begin{bmatrix} r_{1,0} & r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,0} & r_{2,1} & r_{2,2} & r_{2,3} \end{bmatrix} = \begin{bmatrix} 1 & -3/2 & 0 & 1/2 \\ 1 & 0 & -3 & 2 \end{bmatrix}.$$

By Corollary 2, $R_{\ell+2}$ specifies a kernel representation of the system. This means that for any other trajectory of the same system $\bar{w} \in \mathscr{B}_{\bar{T}}$, with $\bar{T} \geq 8$, $R_{\ell+2}\mathscr{H}_{\ell+2}(\bar{w}) = 0$ holds, i.e.,

$$\begin{bmatrix} r_{1,0} & r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,0} & r_{2,1} & r_{2,2} & r_{2,3} \end{bmatrix} \begin{bmatrix} \bar{w}(0) & \bar{w}(1) & \cdots & \bar{w}(\bar{T}-4) \\ \bar{w}(1) & \bar{w}(2) & \cdots & \bar{w}(\bar{T}-3) \\ \bar{w}(2) & \bar{w}(3) & \cdots & \bar{w}(\bar{T}-2) \\ \bar{w}(3) & \bar{w}(4) & \cdots & \bar{w}(\bar{T}-1) \end{bmatrix} = 0. \tag{14}$$

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

Let $\bar{T} = 9$, then one can construct the following matrix:

$$\mathscr{H}_{\ell+3}(\bar{w}) = \begin{bmatrix} \bar{w}(0) & \bar{w}(1) & \cdots & \bar{w}(4) \\ \bar{w}(1) & \bar{w}(2) & \cdots & \bar{w}(5) \\ \bar{w}(2) & \bar{w}(3) & \cdots & \bar{w}(6) \\ \bar{w}(3) & \bar{w}(4) & \cdots & \bar{w}(7) \\ \bar{w}(4) & \bar{w}(5) & \cdots & \bar{w}(8) \end{bmatrix}. \tag{15}$$

Define the following shift-chain-like matrix

$$\Gamma = \begin{bmatrix} r_{1,0} & r_{1,1} & r_{1,2} & r_{1,3} & 0 \\ r_{2,0} & r_{2,1} & r_{2,2} & r_{2,3} & 0 \\ 0 & r_{1,0} & r_{1,1} & r_{1,2} & r_{1,3} \end{bmatrix} \tag{16}$$

which is constructed from $R_{\ell+2}$ by appending each row by $m+p$ zeros (first two rows of Γ), then taking the first p rows of $R_{\ell+2}$ and shifting them by $m+p$ zeros (last row of Γ). Notice that, by (14), it holds that $\Gamma\mathscr{H}_{\ell+3}(\bar{w}) = 0$. Furthermore, if $\text{rank}(\mathscr{H}_{\ell+3}(\bar{w})) = m(\ell+3)+n$, then the rows of Γ represent a basis for the left kernel of $\mathscr{H}_{\ell+3}(\bar{w})$.

In the following, we formalize the procedure followed in Example 3. Recall that, if successful, Algorithm 1 returns a full row rank matrix $R_d \in \mathbb{R}^{pd-n\times qd}$ such that $R_d\mathscr{H}_d(w) = 0$. Given such a matrix, we define, for any $L \geq d \geq \ell+1$, the following shift-chain-like matrix $\Gamma \in \mathbb{R}^{pL-n\times qL}$

$\Gamma =$

$$\begin{bmatrix} r_{1,0} & r_{1,1} & \cdots & r_{1,d-1} & & & & \\ r_{2,0} & r_{2,1} & \cdots & r_{2,d-1} & & & & \\ \vdots & \vdots & \ddots & \vdots & & & & \\ r_{pd-n,0} & r_{pd-n,1} & \cdots & r_{pd-n,d-1} & & & & \\ & r_{1,0} & r_{1,1} & \cdots & r_{1,d-1} & & & \\ & r_{2,0} & r_{2,1} & \cdots & r_{2,d-1} & & & \\ & \vdots & \vdots & \ddots & \vdots & & & \\ & r_{p,0} & r_{p,1} & \cdots & r_{p,d-1} & & & \\ & & \ddots & \ddots & \ddots & \ddots & & \\ & & & r_{1,0} & r_{1,1} & \cdots & r_{1,d-1} & \\ & & & r_{2,0} & r_{2,1} & \cdots & r_{2,d-1} & \\ & & & \vdots & \vdots & \ddots & \vdots & \\ & & & r_{p,0} & r_{p,1} & \cdots & r_{p,d-1} & \end{bmatrix} \tag{17}$$

($L-d$ times)

By construction, the matrix Γ has full row rank, i.e., $\text{rank}(\Gamma) = pL-n$. This is because: 1) the rows of R_d are linearly independent; and 2) the structure of Γ where each row is appended by an appropriate number of zeros.

For any trajectory $\bar{w} \in \mathscr{B}|_{\bar{T}}$ for which $\text{rank}(\mathscr{H}_L(\bar{w})) = mL+n$ (and hence the corresponding $\mathscr{H}_L(\bar{w})$ represents a nonparametric representation of the system as in Corollary 1), the following lemma states that the matrix Γ in (17) constitutes a basis for the left kernel of the Hankel matrix $\mathscr{H}_L(\bar{w})$, i.e., $\Gamma\mathscr{H}_L(\bar{w}) = 0$. Note that $\bar{w}$ need not be measured for the results to hold. However, the existence of such a trajectory is guaranteed by the fact that $\dim(\mathscr{B}|_L) = mL+n$ (see Theorem 1 and [30] for more details).

Lemma 2: Consider an LTI system $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ and let $w \in \mathscr{B}|_T$. Given a full row rank matrix R_d satisfying $R_d\mathscr{H}_d(w) = 0$, then for any $L \geq d \geq \ell+1$ and any other $\bar{w} \in \mathscr{B}|_{\bar{T}}$ satisfying $\text{rank}(\mathscr{H}_L(\bar{w})) = mL+n$, it holds that the rows of the matrix Γ in (17) represent a basis for the left kernel of $\mathscr{H}_L(\bar{w})$, i.e., $\Gamma\mathscr{H}_L(\bar{w}) = 0$ and $\text{rank}(\Gamma) = \dim(\ker(\mathscr{H}_L^\top(\bar{w}))) = pL-n$.

Proof: According to Corollary 2, the matrix R_d specifies a kernel representation of the system. This means that for $\bar{w} \in \mathscr{B}|_{\bar{T}}$, the following holds:

$$R_d\mathscr{H}_d(\bar{w}) = 0. \tag{18}$$

It is assumed that $\text{rank}(\mathscr{H}_L(\bar{w})) = mL+n$ which implies that the dimension of its left kernel is $\dim(\ker(\mathscr{H}_L^\top(\bar{w}))) = pL-n = \text{rank}(\Gamma)$, where the rank of the matrix Γ follows from the discussion below (17). The proof is concluded by noticing that the product of the nonzero elements of Γ with $\mathscr{H}_L(\bar{w})$ involves selected windows of (18) which, along with the remaining zero elements of Γ, results in the zero matrix, i.e., $\Gamma\mathscr{H}_L(\bar{w}) = 0$. Since $\text{rank}(\Gamma) = \dim(\ker(\mathscr{H}_L^\top(\bar{w}))) = pL-n$, the rows of Γ represent a basis for $\ker(\mathscr{H}_L^\top(\bar{w}))$. □

We emphasize that $\bar{w}$ represents some trajectory of the system and *not* available data. As given by Theorem 1, if $\text{rank}(\mathscr{H}_L(\bar{w})) = mL+n$, then $\mathscr{H}_L(\bar{w})$ represents a nonparametric model of the restricted behavior of the system $\mathscr{B}|_L$, i.e., $\text{im}(\mathscr{H}_L(\bar{w})) = \mathscr{B}|_L$. The following theorem is the main result of this section. It shows how one can obtain an alternative nonparametric model of the system. In particular, we show that

$$\text{im}(P) = \mathscr{B}|_L, \qquad \text{where } P = \text{null}(\Gamma). \tag{19}$$

This nonparametric representation has two advantages: 1) it can be obtained from a set of irregular measurements w^- of a potentially (very) small number of data points (that depends on d but is independent of L) as in Algorithm 1; and 2) does not result in an overparameterization of the spanned input–output trajectories (see Corollary 3 below). We point out that a conceptually similar result was reported in [30, eq. ($\mathscr{B}|_L$, KER)] but no proofs were provided to show this result.

Theorem 3: Consider an LTI system $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ and let $\bar{w} \in \mathscr{B}|_{\bar{T}}$. For $L \geq \ell+1$, let $\text{rank}(\mathscr{H}_L(\bar{w})) = mL+n$ and let the rows of Γ denote a basis for its left kernel, i.e., $\Gamma\mathscr{H}_L(\bar{w}) = 0$. Then a matrix of the form $P = \text{null}(\Gamma)$ has $\text{rank}(P) = mL+n$ and satisfies $\text{im}(P) = \text{im}(\mathscr{H}_L(\bar{w}))$, i.e., P represents a nonparametric model of $\mathscr{B}|_L$.

Proof: Let the matrix $\bar{M} \in \mathbb{R}^{qL\times mL+n}$ be formed by $mL+n$ linearly independent columns of $\mathscr{H}_L(\bar{w})$. Clearly

$$\text{im}(\mathscr{H}_L(\bar{w})) = \text{im}(\bar{M}). \tag{20}$$

Note that $mL+n < qL$, which holds since $q = m+p$, $L \geq \ell+1$ by assumption and $p\ell \geq n$. Thus, $\bar{M}$ has a nontrivial left kernel, whose basis is also given by the rows of Γ, i.e., $\Gamma\bar{M} = 0$. Recall that Γ is a full row rank matrix since it constitutes a basis for the left kernel of $\mathscr{H}_L(\bar{w})$. Furthermore, it has a nontrivial right kernel of dimension $qL-(pL-n) = mL+n$, i.e., there exists a matrix $P = \text{null}(\Gamma) \in \mathbb{R}^{qL\times mL+n}$. Notice that the matrix P has full column rank (by definition of a basis) and hence, $\text{rank}(P) = mL+n = \text{rank}(\mathscr{H}_L(\bar{w}))$.

Now, it is left to show that P and $\mathscr{H}_L(\bar{w})$ have the same image. By direct sum of orthogonal subspaces of the matrix $\bar{M}$,

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

the following holds:

$$\begin{aligned} \text{im}\left(\bar{M}\right) \oplus \ker\left(\bar{M}^\top\right) &= \mathbb{R}^{qL} \\ \text{im}\left(\bar{M}\right) \oplus \text{im}\left(\Gamma^\top\right) &= \mathbb{R}^{qL} \end{aligned} \tag{21}$$

where the second equality holds by definition of Γ. This implies that $\text{im}(\bar{M}) = \text{im}(\Gamma^\top)^\perp$. Similarly, the following holds for the matrix Γ

$$\begin{aligned} \text{im}(\Gamma^\top) \oplus \ker(\Gamma) &= \mathbb{R}^{qL} \\ \text{im}(\Gamma^\top) \oplus \text{im}(P) &= \mathbb{R}^{qL} \end{aligned} \tag{22}$$

where the second equality holds by definition of P. This implies that $\text{im}(\Gamma^\top)^\perp = \text{im}(P)$. Since $\text{im}(\bar{M}) = \text{im}(\Gamma^\top)^\perp = \text{im}(P)$, it follows that $\text{im}(\bar{M}) = \text{im}(P)$. This, together with (20) completes the proof. □

Theorem 3 illustrates how a nonparametric model of the restricted behavior, other than $\text{im}(\mathscr{H}_L(\bar{w}))$, can be obtained as the image of an equivalent matrix $P = \text{null}(\Gamma)$, where the rows of Γ represent a basis for $\ker(\mathscr{H}_L^\top(\bar{w}))$. By recalling that Γ can be constructed from R_d as in (17), one can therefore obtain the restricted behavior of $\mathscr{B}|_L$ for any $L \geq d$. It is important to highlight that this allows us to use the available measurements (possibly irregular and of small number of data points) to retrieve the *complete* restricted behavior for *any* length L, using three steps:

i) Using irregularly measured data, compute the kernel representation using, e.g., Algorithm 1.
ii) Build the matrix Γ as in (17).
iii) Obtain $P = \text{null}(\Gamma)$, which according to Theorem 3 satisfies $\text{im}(P) = \mathscr{B}|_L$ for $L \geq d$.

The following corollary follows from Theorem 3 and provides an analogous result to that of Corollary 1 in the case of missing data. In particular, we provide a data-based representation for any complete trajectory of the system $\bar{w}$ as a linear combination of the columns of the matrix P which is constructed by post-processing a sequence of irregularly measured (and potentially short number of) data w^- as in steps (i)–(iii) above.

Corollary 3: Given irregular measurements $w^- \in (\mathbb{R}^q_{\text{ext}})^T$ of a trajectory $w \in \mathscr{B}|_T$ where $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$, suppose Algorithm 1 returns a matrix R_d with $\text{rank}(R_d) = pd - n$. Then, $\bar{w} \in \mathscr{B}|_L$ for any $L \geq d$, if and only if there exists a vector $\beta \in \mathbb{R}^{mL+n}$ such that

$$P\beta = \bar{w} \tag{23}$$

where $P = \text{null}(\Gamma)$ and Γ is built from R_d as in (17).

Similar to Corollary 1 or the fundamental lemma [4], Corollary 3 provides us with a data-based representation of the length-L trajectories of LTI systems (i.e., $\mathscr{B}|_L$) given by the image of the matrix P. However, there are two important differences between Corollary 3 and Corollary 1. 1) Even in the case of complete data, Corollary 3 retrieves $\mathscr{B}|_L$ *for any* $L \geq d$ using a small number of data points which are independent of L, specifically $T \geq (m+1)(d+n) - 1$. In contrast, Corollary 1 (or [4]) requires a number of data points which grows as L grows ($T \geq (m+1)(L+n) - 1$). 2) Unlike the fundamental lemma [4] or Corollary 1, the vector $\beta \in \mathbb{R}^{mL+n}$ is *unique* (for every $\bar{w} \in \mathscr{B}|_L$) since P has full column rank. This is expected since, in order to uniquely define a length-L trajectory of the LTI system, one needs mL parameters for the input and n to fix the initial conditions of the state (cf., [7]). In [33], we recently exploited these two important differences to propose a data-driven predictive control scheme which is both sample and computationally more efficient than existing ones in the literature.

Algorithm 2: Data Completion.

Input: Irregular measurements $w^- \in (\mathbb{R}^q_{\text{ext}})^T$ of a trajectory $w \in \mathscr{B}|_T$ where $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ is of known complexity.
1. Run Algorithm 1 to obtain R_d.
2. For $L = T$, build the matrix Γ as in (17).
3. Obtain $P = \text{null}(\Gamma)$.
4. Solve $P_g\beta = w_g$ for β, then obtain $w_m^- = P_m\beta$.

Output: A completed sequence w where w_m^- is inserted in place of NaN values in w^-.

Recall that Algorithm 1 takes as an input an incomplete sequence of measurements w^- of length T and, if successful, it returns the kernel representation of the system. In the literature of identification from missing data, other techniques rely on first completing the missing data sequence and then computing the kernel representation as the basis of the left null space of the completed Hankel matrix. The former step of data completion is solved by, e.g., structured low-rank data completion which is a nonlinear optimization problem (compare [26] and references therein). Convex relaxations of this problem use the nuclear norm heuristic (cf., [27]). However, these methods are not computationally efficient.

If the data is generated by an LTI system, then Corollary 3 offers an alternative method to solve the data completion problem, upon successful completion of Algorithm 1. In particular, if we denote the given elements of w^- by w_g^- and the missing elements of w^- by w_m^-, then one can use Corollary 3, with $L = T$, to solve $P_g\beta = w_g^-$ for β, where P_g denotes the submatrix of P given by the rows whose indices match the row indices of w_g^-. Notice that the system of equations $P_g\beta = w_g^-$ is consistent and, hence, there always exists a solution β for any w_g^-. Once a solution β is obtained, one can complete the sequence w by solving for $w_m^- = P_m\beta$, where P_m denotes the submatrix of P whose row indices match the row indices of w_m^-. This approach to data completion is summarized in Algorithm 2. In Section VII, we illustrate how Algorithm 2 is computationally more efficient than the nuclear norm heuristic for data completion. However, it hinges on the successful completion of Algorithm 1.

Remark 2: Note that, depending on the pattern of missing data, the outcome of Algorithm 2 may not be unique. However, the completed sequence is unique if and only if P_g has full column rank, specifically $\text{rank}(P_g) = mT + n$. This latter rank condition depends on the pattern of given data.

In the following section, we consider the case of periodically missing outputs with period $\ell + 1$, and impose conditions on the input and the system under consideration such that a kernel

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

representation is guaranteed to be retrieved using Algorithm 1. Combined with the results of Theorem 3 and Corollary 3 above, it is shown that any *complete* finite-length behavior of the system can be retrieved from a set of measurements with periodically missing outputs.

V. Periodically Missing Outputs

In order to guarantee that Algorithm 1 succeeds in returning the kernel representation, we must impose some assumptions on the pattern of the missing data and provide PE-like conditions on the input to the system. In this section, we assume that the system admits a state-space representation as in (3) with (A, B) controllable and a known input–output partitioning of the data $w(t) = \begin{bmatrix} u(t) \\ y(t) \end{bmatrix}$. Output data are assumed to be periodically missing with period $\ell + 1$. For this pattern of missing data, the sequence $w^- \in (\mathbb{R}^q_{\text{ext}})^T$ takes the form

$$w^- = \Big(w(0), \dots, w(\ell-1), \begin{pmatrix} u(\ell) \\ \text{NaN} \end{pmatrix}, w(\ell+1), \dots, w(T-1)\Big). \tag{24}$$

We further assume that $\text{rank}(C) = p < n$ (i.e., $\ell > 1$). This is because when $\ell = 1$, every other measurement will be lost and hence, no two consecutive measurements are available to allow for identification of the system dynamics.

The considered pattern of periodically missing outputs with period $\ell + 1$ is interesting to investigate since a missing data point will appear in every row and every column of the Hankel matrix $\mathscr{H}_d(w^-)$ for $d \geq \ell + 1$. In contrast, for data which are periodically missing with period greater than $\ell + 1$, one can find conditions on T such that the Hankel matrix $\mathscr{H}_{\ell+1}(w^-)$ has $m(\ell + 1) + n$ complete columns. If these columns are linearly independent, then a kernel representation can directly be obtained by taking the left kernel of the submatrix containing the aforementioned linearly independent columns. For data missing with periodicity less than $\ell + 1$, Algorithm 1 does not return a kernel representation of the system. In [34] the so-called lifting operator is employed to facilitate identification of the dynamics on a coarser time scale and later recovers the true system by the reversing the effect of the lift operator. Such a method can handle periodically missing outputs with period less than $\ell + 1$ (including $\ell = 1$), but requires further (more restrictive) assumptions on the system to that end.

We start by illustrating how to systematically choose the submatrices $\bar{H}_i$ in Algorithm 1. Later, we provide conditions on the input to the system such that certain submatrices satisfy a desired rank (a PE-like condition), such that Algorithm 1 is guaranteed to return the kernel representation of the system.

Consider a sequence of irregular measurements $w^- \in (\mathbb{R}^q_{\text{ext}})^T$ where the output portion is periodically missing with period equal to $\ell + 1$. The following two matrices are submatrices of the Hankel matrix $\mathscr{H}_{\ell+2}(w^-)$ (i.e., $d = \ell + 2$)

$$H_1 = \begin{bmatrix} w(0) & w(\ell+1) & \cdots & w((n_c-1)(\ell+1)) \\ \vdots & \vdots & \vdots & \vdots \\ w(\ell-1) & w(2\ell) & \cdots & w((n_c-1)(\ell+1)+\ell-1) \\ \begin{pmatrix} u(\ell) \\ \text{NaN} \end{pmatrix} & \begin{pmatrix} u(2\ell+1) \\ \text{NaN} \end{pmatrix} & \cdots & \begin{pmatrix} u((n_c-1)(\ell+1)+\ell) \\ \text{NaN} \end{pmatrix} \\ w(\ell+1) & w(2\ell+2) & \cdots & w(n_c(\ell+1)) \end{bmatrix}$$

$$H_2 = \begin{bmatrix} w(1) & w(\ell+2) & \cdots & w((n_c-1)(\ell+1)+1) \\ \vdots & \vdots & \vdots & \vdots \\ \begin{pmatrix} u(\ell) \\ \text{NaN} \end{pmatrix} & \begin{pmatrix} u(2\ell+1) \\ \text{NaN} \end{pmatrix} & \cdots & \begin{pmatrix} u((n_c-1)(\ell+1)+\ell) \\ \text{NaN} \end{pmatrix} \\ w(\ell+1) & w(2\ell+2) & \cdots & w(n_c(\ell+1)) \\ w(\ell+2) & w(2\ell+3) & \cdots & w(n_c(\ell+1)+1) \end{bmatrix}. \tag{25}$$

After deleting the NaN block rows, we are left with two matrices $\bar{H}_1, \bar{H}_2 \in \mathbb{R}^{q(\ell+2)-p\times n_c}$, where n_c denotes the number of columns. Lemma 3 below provides conditions on the input such that $\text{rank}(\bar{H}_1) = \text{rank}(\bar{H}_2) = m(\ell + 2) + n$. This rank condition is needed for the submatrices to uncover part of the kernel representation of the system (compare Lemma 1 and Algorithm 1). But first, we make the following assumption which is satisfied, e.g., for systems whose matrix A has only positive real eigenvalues (cf., [35]).

Assumption 1: For a system $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ with $\mathscr{B}_{ss}$ as in (3), the following matrix has full rank:

$$\mathscr{O}_s = \begin{bmatrix} C^\top & (CA)^\top & \cdots & (CA^{\ell-2})^\top & (CA^{\ell+1})^\top \end{bmatrix}^\top \in \mathbb{R}^{p\ell\times n}.$$

Lemma 3: Let Assumption 1 hold and let the input sequence applied to a controllable system $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ satisfy $\text{rank}(\mathcal{H}_{[0,n+\ell+2]}(u)) = m(n + \ell + 3)$ where

$$\mathcal{H}_{[0,n+\ell+2]}(u) := \begin{bmatrix} u_{[0,n+\ell+2]} & \sigma^{\ell+1}u_{[0,n+\ell+2]} & \cdots & \sigma^{(n_c-1)(\ell+1)}u_{[0,n+\ell+2]} \end{bmatrix}. \tag{26}$$

Then the corresponding resulting matrices $\bar{H}_1$ and $\bar{H}_2$ in (25) have $\text{rank}(\bar{H}_1) = \text{rank}(\bar{H}_2) = m(\ell + 2) + n$.

Proof: We will prove the claim for $\bar{H}_1$. Similar steps can be made to show the claim for $\bar{H}_2$. Following a row permutation, the matrix $\bar{H}_1$ can be equivalently written as:

$$\bar{H}_1 = \bar{\Pi}_r \times \left[\begin{array}{cccc} \multicolumn{4}{c}{\mathcal{H}_{[0,\ell+1]}(u)} \\ \hline y(0) & y(\ell+1) & \cdots & y((n_c-1)(\ell+1)) \\ \vdots & \vdots & \vdots & \vdots \\ y(\ell-1) & y(2\ell) & \cdots & y((n_c-1)(\ell+1)+\ell-1) \\ y(\ell+1) & y(2\ell+2) & \cdots & y(n_c(\ell+1)) \end{array}\right]$$

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

where $\bar{\Pi}_r$ is a square full rank permutation matrix of appropriate dimensions. Considering some minimal state-space representation of the system, this can be written as

$$\bar{H}_1 = \bar{\Pi}_r \underbrace{\left[\begin{array}{cccc|c} \multicolumn{4}{c|}{I} & 0 \\ D & 0 & \cdots & 0 & C \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ CA^{\ell-2}B & CA^{\ell-3}B & \cdots & 0 & CA^{\ell-1} \\ CA^{\ell}B & CA^{\ell-1}B & \cdots & D & CA^{\ell+1} \end{array}\right]}_{:=\mathscr{T}_1 \in \mathbb{R}^{m(\ell+2)+p(\ell+1)\times m(\ell+2)+n}} \times \begin{bmatrix} \mathcal{H}_{[0,\ell+1]}(u) \\ X_0 \end{bmatrix}$$

where $X_0 := \begin{bmatrix} x(0) & x(\ell+1) & \cdots & x((n_c-1)(\ell+1)) \end{bmatrix}$. The matrix $\mathscr{T}_1$ has full column rank. This is due to the identity on the upper left block and the first ℓ blocks of the observability matrix on the right lower block.[2] Hence, it holds that $\operatorname{rank}(\bar{H}_1) = \operatorname{rank}\left(\begin{bmatrix} \mathcal{H}_{[0,\ell+1]}(u) \\ X_0 \end{bmatrix}\right)$. Now, it is left to show that $\operatorname{rank}\left(\begin{bmatrix} \mathcal{H}_{[0,\ell+1]}(u) \\ X_0 \end{bmatrix}\right) = m(\ell+2)+n$. Notice that the deeper matrix $\mathcal{H}_{[0,n+\ell+2]}(u)$ has rank $m(n+\ell+3)$ by assumption, which implies that $\mathcal{H}_{[0,n+\ell+1]}(u)$ has rank $m(n+\ell+2)$. This means that the input sequences given by the columns of $\mathcal{H}_{[0,n+\ell+1]}(u)$ are collectively PE (see [6, Def. 2]) of order $n+\ell+2$. Since the system is controllable, it follows from [6, Th. 2.i] that $\operatorname{rank}\left(\begin{bmatrix} \mathcal{H}_{[0,\ell+1]}(u) \\ X_0 \end{bmatrix}\right) = m(\ell+2)+n$. □

Remark 3: Notice that Lemma 3 necessitates that $n_c \geq m(\ell+n+3)$ and hence the length of w^- must be at least $T \geq m(\ell+n+3)(\ell+1)+n+2$.

When the submatrices $\bar{H}_1, \bar{H}_2$ have the desired ranks as in the above lemma, their left kernels (the bases of which are given by full row rank matrices $\bar{Z}_i$ satisfying $\bar{Z}_i\bar{H}_i = 0$) have the following dimensions

$$\begin{aligned} \dim(\ker(\bar{H}_i^\top)) &= \#\text{rows}(\bar{H}_i) - \operatorname{rank}(\bar{H}_i) \\ &= q(\ell+2) - p - m(\ell+2) - n \\ &= p(\ell+1) - n. \end{aligned}$$

By inserting zero columns in the location of deleted block rows of the corresponding H_i, one can extend $\bar{Z}_i$ to Z_i which then satisfies $Z_i\mathscr{H}_{\ell+2}(w) = 0$, for $i = 1, 2$ (compare Example 1 and Lemma 1). Concatenating the two, we get

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \mathscr{H}_{\ell+2}(w) = \mathscr{Z}\mathscr{H}_{\ell+2}(w) = 0$$

where $\mathscr{Z} \in \mathbb{R}^{2p(\ell+1)-2n\times q(\ell+2)}$. The following theorem shows that $\mathscr{Z}$ contains $p(\ell+2)-n$ linearly independent rows that specify a kernel representation of the system.

[2]When showing the claim for $\bar{H}_2$, the bottom right block of $\mathscr{T}_2$ contains instead a sample-based observability matrix which is assumed to have full column rank by Assumption 1.

Theorem 4: Consider a controllable system $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ as in (3) with $\operatorname{rank}(C) < n$ and let Assumption 1 hold. Given irregular measurements $w^- \in (\mathbb{R}^q_{\text{ext}})^T$ of the form (24) of a trajectory $w \in \mathscr{B}|_T$, let $\operatorname{rank}(\mathcal{H}_{[0,n+\ell+2]}(u)) = m(n+\ell+3)$ where $\mathcal{H}_{[0,n+\ell+2]}(u)$ is as in (26). Then, Algorithm 1 successfully returns a matrix $R_{\ell+2}$ which specifies a kernel representation of the system and is given by $p(\ell+2)-n$ linearly independent rows of the matrix $\mathscr{Z}$.

Proof: See Appendix. □

Combining the results of Corollary 3 and Theorem 4, we obtain the following corollary, which is analogous to [4, Th. 1] for the case of periodically missing data. In particular, we provide conditions on the input such that any *complete* finite-length behavior of the system can be obtained. Unlike [4], we only have periodically missing data and the resulting representation of all input/output trajectories is not overparametrized (see the discussion following Corollary 3).

Corollary 4: Consider a controllable system $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ and let the conditions of Theorem 4 hold. Then, for $L \geq \ell+2$, $\bar{w} \in \mathscr{B}|_L$ if and only if there exists $\beta \in \mathbb{R}^{mL+n}$ such that

$$P\beta = \bar{w} \tag{27}$$

where $P = \operatorname{null}(\Gamma)$ and Γ is given as in (17).

So far, we only considered exact (i.e., noise-free) data. When the data are noisy, then the rank condition in (GPE) is, in general, violated. This necessitates certain modifications of our results in Algorithm 1. In the following section, we consider the case of noisy and irregularly measured data, and discuss the effect of noise on the results of Algorithm 1.

VI. Handling Noisy Data

In this section, we consider (irregular) measurements which are affected by noise. In particular, one has access to $\widetilde{w}^-$ whose (available) elements take the form $\widetilde{w}_k = w_k + \varepsilon_k$ (errors-in-variables setting), where ε denotes some noise realization, e.g., sampled from a Gaussian or uniform distribution. It is important to note that, even when the available data are complete, the presence of noise, in general, results in the violation of the rank condition (GPE). In fact, additive uniform random noise or Gaussian noise will almost surely increase the rank i.e., $\operatorname{rank}(\mathscr{H}_L(\widetilde{w})) \geq mL+n$. Such a rank condition, however, is crucial for obtaining data-based representations of LTI systems (cf., Theorem 1). In this case, a common practice is to obtain a low rank approximation $\widehat{\mathscr{H}_L}(\widetilde{w})$ of $\mathscr{H}_L(\widetilde{w})$ with $\operatorname{rank}(\widehat{\mathscr{H}_L}(\widetilde{w})) = mL+n$ and then use it as an estimate of the true finite-length behavior of the LTI system that generated the data, i.e., $\widehat{\mathscr{B}}|_L := \operatorname{im}(\widehat{\mathscr{H}_L}(\widetilde{w}))$ (cf. [36]). Low rank approximation can be done using, e.g., truncated singular-value decomposition (TSVD, cf. [37]), or using structured low-rank approximation of Hankel matrices (SLRA, cf. [36]).

In Section III-B, we addressed the noise-free setting by proposing Algorithm 1 which, if successful, returns a basis for the left kernel of a Hankel matrix with missing entries. An important step in Algorithm 1 involved selecting submatrices (denoted $\bar{H}_i$) with specific rank. Since, in general, the presence of noise will result in full rank submatrices $\bar{H}_i$ in Algorithm 1, it

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

is not clear anymore how to proceed in order to retrieve a basis for the left kernel. To that end, we propose modifying Algorithm 1 by introducing an intermediate low rank approximation step. In particular, one would now search over submatrices $\bar{H}_i$ with $\text{rank}(\bar{H}_i) \geq md + n$ then apply TSVD to obtain an approximate matrix, denoted $\widehat{H}_i$, such that $\text{rank}(\widehat{H}_i) = md + n$. The rest of the procedure continues as in Algorithm 1 with the main difference that the exit condition on the rank of $\mathscr{Z}$ is an inequality (rather than an equality). This is because $\mathscr{Z}$ now contains vectors that do not necessarily represent annihilators and, hence, $\mathscr{Z}$ can have larger rank (i.e., corresponding to a system of different complexity than the true system). To overcome this, we introduce another low rank approximation step on $\mathscr{Z}$ and obtain a matrix $\widehat{\mathscr{Z}}$ such that $\text{rank}(\widehat{\mathscr{Z}}) = pd - n$. As a result, Algorithm 3, if successful,[3] returns a matrix $\widehat{R}_d$ (given by $pd - n$ linearly independent rows of $\widehat{\mathscr{Z}}$) which specifies an *estimate* of the true kernel representation specified by (the unknown) R_d. The modification of Algorithm 1 to account for noisy data is summarized in Algorithm 3.

We conclude this section with some remarks on uncertainty quantification in the behavioral framework (see, e.g., [38] and [39]). A good measure for quantifying the quality of the estimate $\ker(\widehat{R}_d) =: \widehat{\mathscr{B}}|_d$ is given by the principal angles [40] between the true and estimated subspaces, i.e., $\mathscr{B}|_d$ and $\widehat{\mathscr{B}}|_d$, respectively. Specifically, two subspaces are said to be close to each other if the largest principle angle $\theta_{\max}(\mathscr{B}|_d, \widehat{\mathscr{B}}|_d)$, which depends on the noise, is small. In the last step of Algorithm 3, we introduced a low-rank approximation step on the matrix $\mathscr{Z}$ which ensures that the estimated kernel representation corresponds to a system with the same complexity as that of the true system. Therefore, it directly follows that having $\theta_{\max}(\mathscr{B}|_d, \widehat{\mathscr{B}}|_d) = 0$ is equivalent to $\mathscr{B}|_d = \widehat{\mathscr{B}}|_d$. Our future work will focus on providing quantitative error bounds on the angle $\theta_{\max}(\mathscr{B}|_d, \widehat{\mathscr{B}}|_d)$ in terms of the noise bound.

In the next section, we illustrate the results of this article with different simulation examples.

VII. Simulation Examples

In this section, we will first empirically investigate the success/failure of Algorithm 1 in the case of random patterns of missing input and/or output instances. Then, we compare the computational efficiency of solving a data completion problem using the results of this article (see discussion after Corollary 3) against the nuclear norm heuristic [27]. Finally, we illustrate the results of Algorithm 3 on a (simplified) physiological model of cardiovascular hemodynamics that describes the effect of fluid infusion on the change of blood volume (cf., [19]). All simulations were done on a standard Intel i7-10875H (2.30 GHz) machine with 16 GB of memory.

[3]Similar to the discussion below Algorithm 1, Algorithm 3 is deemed successful if it terminates with $\text{rank}(\widehat{\mathscr{Z}}) = pd - n$.

Algorithm 3: Computing a Kernel Representation Using Noisy Data.

Input: Noisy and irregular measurements $\widetilde{w}^- \in (\mathbb{R}^q_{\text{ext}})^T$ of a trajectory $w \in \mathscr{B}|_T$ where $\mathscr{B} \in \partial\mathscr{L}^q_{m,n,\ell}$ is of known complexity.

1. Construct the extended sequence $\widetilde{w}_{\text{ext}}$ by appending $\widetilde{w}^-$ with T instances of NaN.
2. Construct the matrix $\mathscr{H}_T(\widetilde{w}_{\text{ext}})$, and initialize $\mathscr{Z} = [\,]$.
3. **For** $\ell + 1 \leq d \leq T$ **do**
   - Search over all submatrices of $\mathscr{H}_d(\widetilde{w}^-)$, i.e., $\bar{H}_i \in \mathbb{R}^{\eta_i \times \rho_i}$ that have $\text{rank}(\bar{H}_i) \geq md + n$.
   - Compute a low rank approximation of $\bar{H}_i$, denoted $\widehat{H}_i$, such that $\text{rank}(\widehat{H}_i) = md + n$.
   - Compute a basis for $\ker(\widehat{H}_i^\top)$, i.e., a full row rank matrix $\widehat{Z}_i \in \mathbb{R}^{\zeta_i \times \eta_i}$ such that $\widehat{Z}_i \widehat{H}_i = 0$.
   - Extend $\widehat{Z}_i$ to Z_i by inserting zero columns at the location of the rows removed from $\mathscr{H}_d(\widetilde{w}^-)$ to obtain $\widehat{H}_i$.
   - Concatenate all the above matrices Z_i in $\mathscr{Z}$, i.e., $\mathscr{Z} := \begin{bmatrix}[\mathscr{Z} & \mathbf{0}]^\top & Z_1^\top & Z_2^\top & \cdots\end{bmatrix}^\top$, where $\mathbf{0}$ denotes a matrix of zeros of appropriate dimensions.
   - **If** $\text{rank}(\mathscr{Z}) \geq pd - n$, **break**

   end for
4. **If** $\text{rank}(\mathscr{Z}) = pd - n$, **set** $\widehat{\mathscr{Z}} = \mathscr{Z}$.
   Else, compute a low rank approximation of $\mathscr{Z}$, denoted $\widehat{\mathscr{Z}}$, such that $\text{rank}(\widehat{\mathscr{Z}}) = pd - n$.

Output: The matrix $\widehat{R}_d$, specifying an approximate kernel representation of the system, is given by any $pd - n$ linearly independent rows of $\widehat{\mathscr{Z}}$.

A. Performance of Algorithm 1 Under Random Patterns of Missing Data

In this section, we empirically evaluate the performance of Algorithm 1 when the data are missing randomly. We generate 100 random discrete-time LTI systems with $(m, p, n, \ell) = (0, 1, 2, 2)$. For each system, we generate a trajectory w of length $T_{\max} = 200$ by simulating the response of each system to random initial conditions sampled from a uniform distribution $(U(0, 1))^2$. To investigate the performance of Algorithm 1 in identifying each system, we perform several experiments using irregular measurements $w^- \in (\mathbb{R}^q_{\text{ext}})^T$ of the above trajectory w. For each experiment, we vary the length of $w^- \in (\mathbb{R}^q_{\text{ext}})^T$ that is passed to the algorithm from $T = 20$ to $T = T_{\max} = 200$ as well as the percentage of given to missing data from 10% up to 100%.

For each system and percentage of missing data, the following was repeated 100 times: 1) the location of the missing samples was randomly generated using the MATLAB function `randperm`, then 2) Algorithm 1 was run and checked for success (assigned a value of one) or failure (assigned a value of zero). We additionally validated successful outcomes of Algorithm 1

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

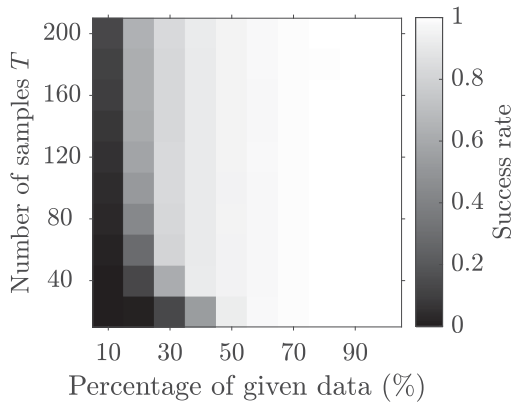


Fig. 1. Transition diagram from failure (black) to success (white) when the data are missing randomly. It can be seen that the success rate of Algorithm 1 increases as both the length of the sequence T and the fraction of given to missing samples increases.

TABLE I
COMPARISON: DATA COMPLETION PROBLEM

Compared values		Algorithm 2	Nuclear norm
$T = 200$	Time (s)	0.0932	57.9953
	Error	6.3278×10^{-14}	9.3051×10^{-9}
$T = 500$	Time (s)	0.2258	N/A
	Error	6.1301×10^{-14}	N/A

by comparing the output of the algorithm to the kernel representation computed from the Hankel matrix of the complete data of the same depth. The average time it took to complete one run of Algorithm 1 was 0.0381 s. Finally, for each percentage of missing data, the results were averaged over all 100 runs and all 100 systems. Fig. 1 shows a gray-scale plot of the results. It can be seen how a transition occurs from failure to success (from dark to light) as the length of the sequence T and the fraction of given samples increase. Notice that the algorithm consistently succeeded when less than 30% of the data are missing. One possible explanation is that the random pattern of missing samples (which was obtained using MATLAB's `randperm` command, see above) did not result in pathological cases where the algorithm fails to find a kernel representation.

B. Data Completion: Comparison Versus Nuclear Norm Heuristic

Following Corollary 3, we briefly discussed how one can use the results of this article to solve the data completion problem (cf., Algorithm 2). The goal of this example is to compare the computational performance of our proposed method in Algorithm 2 against the nuclear norm heuristic [27] when solving the data completion problem.

For the test system, we use the one in [28] which is a single-output six-dimensional slightly damped autonomous system. A trajectory w of length $T = 500$ was generated by simulating the response of the system when all states are initialized from one. For this example, a total of 276 samples were periodically missing (with different periodicities). Table I illustrates the results of the data completion problem. The error is measured in terms of the 2−norm, i.e., $e = \|w - \hat{w}\|_2 / \|w\|_2$, where $\hat{w}$ is the sequence returned by the two methods. When only the first 200 samples of w^- are used, both approaches manage to solve the data completion problem up to numerical accuracy (in this case, the ratio of missing data is around 33%). However it is clear that our proposed method in Algorithm 2 is much faster than the nuclear norm minimization approach. When using the entire sequence of length $T = 500$, the nuclear norm approach failed to return a result (due to insufficient memory). This is caused by the increasingly growing computations that are required to solve an optimization problem (here, we used the default solver in CVX [41]). In contrast, Algorithm 2 only uses linear algebra operations, which solve the data completion problem with (fast) computation time even when the entire sequence is used.

C. Case Study: Physiological Systems

In this example, we consider a (simplified LTI) physiological model of cardiovascular hemodynamics that describes the effect of fluid infusion on the change of blood volume (cf., [19]). In continuous time, this takes the form

$$\frac{d}{dt}\begin{bmatrix}\Delta V_B(t)\\ \breve{u}(t)\end{bmatrix} = \begin{bmatrix}-K & \frac{K}{1+\alpha}\\ 0 & 0\end{bmatrix}\begin{bmatrix}\Delta V_B(t)\\ \breve{u}(t)\end{bmatrix} + \begin{bmatrix}1\\ \frac{1}{1+\alpha}\end{bmatrix} u(t)$$
$$y(t) = \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}\Delta V_B(t)\\ \breve{u}(t)\end{bmatrix} \tag{28}$$

where $\Delta V_B(t)$ is the change in blood volume, $u(t), \breve{u}(t)$ is the rate of fluid infusion and the total infused volume, respectively. The parameter α specifies the steady-state split of fluid gain between the intravascular and extravascular compartments, while the parameter K specifies the intercompartmental rate of fluid shift. For details, the reader is referred to [19]. Even in critical care settings, the (change of) blood volume is typically difficult to measure consecutively and, instead, is estimated based on blood hematocrit levels [42], [43]. Furthermore, the patient-specific parameters α, K are unknown and differ from one patient to another. Therefore, it is convenient in such a setting that one applies our proposed data-driven methods, which are suited for handling missing data, in order to recover the finite-length behavior of such unknown dynamical systems.

This example aims to illustrate the usefulness of our results when blood volume measurements are only irregularly measured and are further affected by measurement noise. Specifically, we will use Algorithm 3 to retrieve an approximate kernel representation of the system, which is later used to retrieve an estimate of the complete finite length behavior of the system as well as an estimate of the missing samples. After discretizing the model (28) with a sampling time of $T_s = 1$ min, we run an open loop simulation where fluid is administered in boluses (i.e., not random inputs, see Fig. 2), and collect corresponding input–output data of length $T = 150$ min. The (unknown) parameters used are $K = 0.5, \alpha = 1.3$. We consider the case where 40% of the output samples are (randomly) missing, while the remaining samples are affected by noise $\varepsilon \in \mathbb{R}^{pT}$ sampled from a zero-mean Gaussian distribution scaled to different noise

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

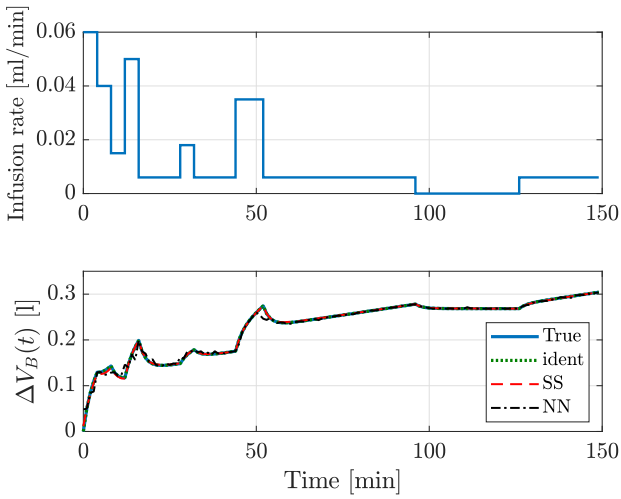


Fig. 2. Illustration of the true versus estimated trajectories using the three methods: structured low-rank approximation method (`ident`), our proposed subspace method (SS), and the nuclear norm method (NN). Unlike `ident`, SS does not require prior model knowledge or initialization steps but still performs similarly to `ident`.

TABLE II
COMPARISON: APPROXIMATE DATA COMPLETION

γ	`ident`	SS	NN
0	0	0	0.0284
2×10^{-4}	0.0001	0.0014	0.0284
4×10^{-4}	0.0001	0.0046	0.0284
6×10^{-4}	0.0002	0.0041	0.0286
8×10^{-4}	0.0003	0.0018	0.0284
10^{-3}	0.0003	0.0072	0.0283

levels, i.e., $\widetilde{y} = y + \gamma \frac{\|y\|}{\|\varepsilon\|}\varepsilon$ for $\gamma \in \{0, 2 \times 10^{-4}, 4 \times 10^{-4}, 6 \times 10^{-4}, 8 \times 10^{-4}, 10^{-3}\}$.

For each value of γ, we run Algorithm 3 to retrieve an approximate kernel representation of the system $\widehat{R}_d$, which is later used to retrieve an estimate of the complete finite length behavior of the system of length equal to T (see steps (i)–(iii) before Corollary 3). This is later used to estimate the missing output samples in the original sequence (using similar steps as in Algorithm 2). We refer to this as the subspace method (SS) and compare its performance against the nuclear norm heuristic (NN). Another method to solve the approximate data completion problem is using the `ident` method for identification in the behavioral setting [44], which is based on structured low-rank approximation [36]. This latter method requires prior model knowledge to initialize the algorithm, which we set to be the exact true model.

Table II contains a summary of the estimation error measured in terms of the 2−norm, i.e., $e = \|w - \hat{w}\|_2 / \|w\|_2$, where $\hat{w}$ is the sequence returned by the three methods. It is important to point out that the `ident` method computes the maximum likelihood estimator for the true behavior (see [36, Prop. 6]). Since, moreover, it is initialized at the true model, `ident` returns the theoretically best possible result using the available data and, hence, is viewed as the benchmark against which SS and NN methods are compared. It can be seen that our direct method (SS) returns very close results to that of `ident`, without using any prior knowledge about the system. This illustrates the advantages of Algorithm 3 which does not use any prior model knowledge (except known complexity), but still manages to return good results compared to `ident`. For the nuclear norm method (NN), we notice the following. 1) NN performs worse in the presence of noise, and fails to complete the sequence even when there is no noise in the data. This is not too surprising, as the nuclear norm method is a heuristic convex relaxation of the exact (nonlinear) data completion problem. 2) Although NN performs worse than `ident` and SS, it is more robust (i.e., errors do not vary depending on noise level). This is potentially due to the fact that NN does not impose any structure/rank constraint on the matrix and returns the best fit of the data by minimizing the nuclear norm of the Hankel matrix. Since the noise is sampled from a zero mean distribution, small levels of noise γ did not largely affect the resulting fitting errors. Fig. 2 shows a representative result for the true versus estimated trajectory using the three different methods when $\gamma = 10^{-3}$.

VIII. CONCLUSION

In this article, we investigated the problem of obtaining nonparametric data-based system representations from a set of irregularly measured data. For the general case of multi-input multi-output systems and random patterns of missing data, we improved an existing algorithm from the literature which now allows for computing the (not necessarily minimal) kernel representation of the system from irregularly measured data. Assuming that the system's complexity is known, one can check if the output of the algorithm corresponds to the kernel representation upon running the algorithm.

If successful, we showed how one can use the output of the algorithm to obtain any *complete* finite-length behavior of the system by exploiting the properties of the Hankel matrix structure. This result has important properties, namely that the complete behavior of any length can be retrieved independent of the number of available (irregularly measured) values. For the case of periodically missing outputs with period $\ell + 1$, we provided conditions on the input, such that any *complete* finite-length behavior of the system can be obtained.

To handle noisy and irregularly measured data, we proposed a modification of our results by introducing low rank approximation steps in the above mentioned algorithm. If successful, the algorithm returns an estimate of kernel representation of the system than can be later used to retrieve an estimate of the finite-length behavior. The accuracy of such an estimate depends on the noise levels in the data.

The results of this article can be further extended in several directions. For instance, the guarantees provided in Section V considered the case where the output portion of the input/output data is missing. It was observed in simulations that our proposed methods work just as well when any p instances of the data sample $w(t)$ were missing, whether it is the entire output portion or a combination of input and output portions. An interesting question for future research is how to impose conditions on the available inputs (rather than assuming availability of all inputs)

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

such that guarantees similar to those in Section V can be shown. Studying other patterns of missing data, apart from periodically missing ones as in Section V is also an interesting problem for future work. Furthermore, in this work we have assumed that the complexity of the system which generated the data is known. Determining a system's complexity from (noisy) input–output data is an open problem and is an interesting topic for future research. Finally, rigorous investigation of the effect of noise in the data and quantifying the uncertainty between the true and estimated finite length behaviors is an ongoing research work.

APPENDIX

A. Proof of Theorem 4

Notice that $\mathscr{Z} \in \mathbb{R}^{2p(\ell+1)-2n \times q(\ell+2)}$ has $2p(\ell+1)-2n \geq p(\ell+2)-n$ rows. Recall that if Algorithm 1 is successful with $d=\ell+2$, the kernel representation of the system would then be given by $p(\ell+2)-n$ linearly independent rows of $\mathscr{Z}$. Therefore, our goal now is to prove that $\operatorname{rank}(\mathscr{Z}) = \dim(\ker(\mathscr{H}_{\ell+2}^{\top}(w))) = p(\ell+2)-n$.

Consider again the matrix $\bar{H}_2$ (with the second to last block row expanded by partitioning w appropriately)

$$\bar{H}_2 = \begin{bmatrix} w(1) & w(\ell+2) & \cdots & w((n_c-1)(\ell+1)+1) \\ w(2) & w(\ell+3) & \cdots & w((n_c-1)(\ell+1)+2) \\ \vdots & \vdots & \vdots & \vdots \\ u(\ell) & u(2\ell+1) & \cdots & u((n_c-1)(\ell+1)+\ell) \\ \begin{pmatrix} u(\ell+1) \\ y(\ell+1) \end{pmatrix} & \begin{pmatrix} u(2\ell+2) \\ y(2\ell+2) \end{pmatrix} & \cdots & \begin{pmatrix} u(n_c(\ell+1)) \\ y(n_c(\ell+1)) \end{pmatrix} \\ w(\ell+2) & w(2\ell+3) & \cdots & w(n_c(\ell+1)+1) \end{bmatrix}$$

and recall that $\operatorname{rank}(\bar{H}_2) = m(\ell+2)+n$ due to Lemma 3. If the block row $\begin{bmatrix} y(\ell+1) & y(2\ell+2) & \cdots & y(n_c(\ell+1)) \end{bmatrix}$ is removed from the above matrix, we obtain

$$\tilde{H} = \begin{bmatrix} w(1) & w(\ell+2) & \cdots & w((n_c-1)(\ell+1)+1) \\ w(2) & w(\ell+3) & \cdots & w((n_c-1)(\ell+1)+2) \\ \vdots & \vdots & \vdots & \vdots \\ u(\ell) & u(2\ell+1) & \cdots & u((n_c-1)(\ell+1)+\ell) \\ u(\ell+1) & u(2\ell+2) & \cdots & u((n_c-1)(\ell+1)+\ell+1) \\ w(\ell+2) & w(2\ell+3) & \cdots & w(n_c(\ell+1)+1) \end{bmatrix}$$

where $\tilde{H} \in \mathbb{R}^{m(\ell+2)+p\ell \times n_c}$. Premultiplying by a square full rank permutation matrix $\tilde{\Pi}_r$, $\tilde{H}$ can be expressed as

$$\tilde{H} = \tilde{\Pi}_r \times \left[\begin{array}{c} \mathcal{H}_{[1,\ell+2]}(u) \\ \hline \begin{matrix} y(1) & y(\ell+2) & \cdots & y((n_c-1)(\ell+1)+1) \\ y(2) & y(\ell+3) & \cdots & y((n_c-1)(\ell+1)+2) \\ \vdots & \vdots & \vdots & \vdots \\ y(\ell-1) & y(2\ell) & \cdots & y((n_c-1)(\ell+1)+\ell-1) \\ y(\ell+2) & y(2\ell+3) & \cdots & y(n_c(\ell+1)+1) \end{matrix} \end{array}\right]$$

$$= \tilde{\Pi}_r \underbrace{\left[\begin{array}{ccccc} \multicolumn{4}{c}{I} & 0 \\ \hline D & 0 & \cdots & 0 & C \\ CB & D & \cdots & 0 & CA \\ \vdots & \vdots & \vdots & & \vdots \\ CA^{\ell-3}B & CA^{\ell-4}B & \cdots & 0 & CA^{\ell-2} \\ CA^{\ell}B & CA^{\ell-1}B & \cdots & D & CA^{\ell+1} \end{array}\right]}_{:=\mathscr{T}_3 \in \mathbb{R}^{m(\ell+2)+p\ell \times m(\ell+2)+n}} \begin{bmatrix} \mathcal{H}_{[1,\ell+2]}(u) \\ X_1 \end{bmatrix}$$

where $X_1 := \begin{bmatrix} x(1) & x(\ell+2) & \cdots & x((n_c-1)(\ell+1)+1) \end{bmatrix}$ and the matrix $\mathscr{T}_3$ is square or has more rows than columns (since $p\ell \geq n$). Further, $\mathscr{T}_3$ has full column rank due to the identity on the upper left block and the full rank matrix on the lower right block (see Assumption 1). Following the same arguments as in Lemma 3, we find that $\operatorname{rank}(\tilde{H}) = m(\ell+2)+n = \operatorname{rank}(\bar{H}_2)$.

This implies that the deleted block row, i.e., $\begin{bmatrix} y(\ell+1) & y(2\ell+2) & \cdots & y(n_c(\ell+1)) \end{bmatrix}$, can be expressed as a linear combination of the other rows of $\bar{H}_2$. In other words, we can always select p rows in $\bar{Z}_2$ (without loss of generality, let them be the first p rows) such that $\bar{Z}_2$ takes the form in (29) shown at the bottom of this page, and satisfies $\bar{Z}_2 \bar{H}_2 = 0$, where

$$\bar{Z}_2 = \begin{bmatrix} z^{(2)}_{[1,p],0} & \cdots & \bar{\zeta}^{(2)}_{[1,p],\ell-1} & \bar{\zeta}^{(2)}_{[1,p],\ell} & I_p & z^{(2)}_{[1,p],\ell+1} \\ z^{(2)}_{[p+1,p(\ell+1)-n],0} & \cdots & \bar{\zeta}^{(2)}_{[p+1,p(\ell+1)-n],\ell-1} & \bar{\zeta}^{(2)}_{[p+1,p(\ell+1)-n],\ell} & \underline{\zeta}^{(2)}_{[p+1,p(\ell+1)-n],\ell} & z^{(2)}_{[p+1,p(\ell+1)-n],\ell+1} \end{bmatrix} \tag{29}$$

$$\mathscr{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \left[\begin{array}{cccccccc} z^{(1)}_{[1,p(\ell+1)-n],0} & \cdots & \bar{\zeta}^{(1)}_{[1,p(\ell+1)-n],\ell-1} & \underline{\zeta}^{(1)}_{[1,p(\ell+1)-n],\ell-1} & \bar{\zeta}^{(1)}_{[1,p(\ell+1)-n],\ell} & 0_{p(\ell+1)-n \times p} & z^{(1)}_{[1,p(\ell+1)-n],\ell+1} \\ \hline z^{(2)}_{[1,p],0} & \cdots & \bar{\zeta}^{(2)}_{[1,p],\ell-1} & 0_{p \times p} & \bar{\zeta}^{(2)}_{[1,p],\ell} & I_p & z^{(2)}_{[1,p],\ell+1} \\ z^{(2)}_{[p+1,p(\ell+1)-n],0} & \cdots & \bar{\zeta}^{(2)}_{[p+1,p(\ell+1)-n],\ell-1} & 0_{p\ell-n \times p} & \bar{\zeta}^{(2)}_{[p+1,p(\ell+1)-n],\ell} & \underline{\zeta}^{(2)}_{[p+1,p(\ell+1)-n],\ell} & z^{(2)}_{[p+1,p(\ell+1)-n],\ell+1} \end{array}\right]. \tag{30}$$

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.

$z_{i,j}^{(2)} = \begin{bmatrix} \bar{\zeta}_{i,j}^{(2)} & \underline{\zeta}_{i,j}^{(2)} \end{bmatrix}$ and $\bar{\zeta}_{i,j}^{(2)} \in \mathbb{R}^{1\times m}$ while $\underline{\zeta}_{i,j}^{(2)} \in \mathbb{R}^{1\times p}$ (the same notation is used later for $\bar{Z}_1$ in (30) shown at the bottom of the previous page).

Since $\text{rank}(\bar{H}_2) = \text{rank}(\mathscr{H}_{\ell+2}(w)) = m(\ell+2)+n$, $\bar{Z}_2$ can be extended by inserting zeros corresponding to the location of NaN entries in H_2 such that the matrix $\mathscr{Z}$ takes the form in (30) and satisfies $\mathscr{Z}\mathscr{H}_{\ell+2}(w) = 0$ (see Lemma 1). Notice that Z_1 has full row rank, since $\bar{Z}_1$ has full row rank by definition of a basis for the left kernel of $\bar{H}_1$. Moreover, the first p rows in Z_2 are linearly independent from the rows of Z_1, due to the location of the identity matrix I_p in (30). Therefore, it holds that $\text{rank}(\mathscr{Z}) = p(\ell+1) - n + p = p(\ell+2) - n$, which completes the proof.

References

[1] W. T. Miller, R. S. Sutton, and P. J. Werbos, *Neural Networks for Control*. Cambridge, MA, USA: MIT Press, 1995.

[2] T. Beckers and S. Hirche, "Stability of Gaussian process state space models," in *Proc. Eur. Control Conf.*, 2016, pp. 2275–2281.

[3] J. C. Willems, "From time series to linear system–part I: Finite dimensional linear time invariant systems, part II: Exact modelling, part III: Approximate modelling.," *Automatica*, vol. 22, no. 5, pp. 561–580, 1986.

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

[5] 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.

[6] 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. 062–607, Jul. 2020.

[7] J. Berberich and F. Allgöwer, "A trajectory-based framework for data-driven system analysis and control," in *Proc. Eur. Control Conf.*, 2020, pp. 1365–1370.

[8] G. Pan, R. Ou, and T. Faulwasser, "On a stochastic fundamental lemma and its use for data-driven optimal control," *IEEE Trans. Autom. Control*, vol. 68, no. 10, pp. 5922–5937, Oct. 2023.

[9] C. Verhoek, R. Tóth, S. Haesaert, and A. Koch, "Fundamental lemma for data-driven analysis of linear parameter-varying systems," in *Proc. 60th IEEE Conf. Decis. Control*, 2021, pp. 5040–5046.

[10] J. G. Rueda-Escobedo and J. Schiffer, "Data-driven internal model control of second-order discrete volterra systems," in *Proc. 59th IEEE Conf. Decis. Control*, 2020, pp. 4572–4579.

[11] M. Alsalti, V. G. Lopez, J. Berberich, F. Allgöwer, and M. A. Müller, "Data-based control of feedback linearizable systems," *IEEE Trans. Autom. Control*, vol. 68, no. 11, pp. 7014–7021, Nov. 2023.

[12] I. Markovsky and F. Dörfler, "Behavioral systems theory in data-driven analysis, signal processing, and control," *Annu. Rev. Control*, vol. 52, pp. 42–64, 2021.

[13] I. Markovsky, L. Huang, and F. Dörfler, "Data-driven control based on the behavioral approach: From theory to applications in power systems," *IEEE Control Syst. Mag.*, vol. 43, no. 5, pp. 28–68, Oct. 2023.

[14] L. Huang, J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control for grid-connected power converters," in *Proc. IEEE 58th Conf. Decis. Control*, 2019, pp. 8130–8135.

[15] M. Yin, H. Cai, A. Gattiglio, F. Khayatian, R. S. Smith, and P. Heer, "Data-driven predictive control for demand side management: Theoretical and experimental results," *Appl. Energy*, vol. 353, 2024, Art. no. 122101.

[16] V. Chinde, Y. Lin, and M. J. Ellis, "Data-enabled predictive control for building HVAC systems," *J. Dyn. Syst., Meas., Control*, vol. 144, no. 8, 2022, Art. no. 081001.

[17] E. Elokda, J. Coulson, P. N. Beuchat, J. Lygeros, and F. Dörfler, "Data-enabled predictive control for quadcopters," *Int. J. Robust Nonlinear Control*, vol. 31, no. 18, pp. 8916–8936, 2021.

[18] B. Azimi-Sadjadi, "Stability of networked control systems in the presence of packet losses," in *Proc. 42nd IEEE Int. Conf. Decis. Control*, 2003, pp. 676–681.

[19] R. Bighamian, A. T. Reisner, and J.-O. Hahn, "A lumped-parameter subject-specific model of blood volume response to fluid infusion," *Front. Physiol.*, vol. 7, 2016, Art. no. 390.

[20] J. Berberich, J. W. Dietrich, R. Hoermann, and M. A. Müller, "Mathematical modeling of the pituitary–thyroid feedback loop: Role of a TSH-T3-shunt and sensitivity analysis," *Front. Endocrinol.*, vol. 9, 2018, Art. no. 91.

[21] S. Wildhagen, J. Berberich, M. Hertneck, and F. Allgöwer, "Data-driven analysis and controller design for discrete-time systems under aperiodic sampling," *IEEE Trans. Autom. Control*, vol. 68, no. 6, pp. 3210–3225, Jun. 2023.

[22] I. Markovsky and F. Dörfler, "Data-driven dynamic interpolation and approximation," *Automatica*, vol. 135, 2022, Art. no. 110008.

[23] G. Heinig and K. Rost, *Algebraic Methods for Toeplitz-Like Matrices and Operators*. Berlin, Germany: De Gruyter, 1984.

[24] G. Heinig and P. Jankowski, "Kernel structure of block Hankel and Toeplitz matrices and partial realization," *Linear Algebra Appl.*, vol. 175, pp. 1–30, 1992.

[25] R. Pintelon and J. Schoukens, "Frequency domain system identification with missing data," in *Proc. 37th IEEE Conf. Decis. Control*, 1998, pp. 701–705.ptpt

[26] I. Markovsky and K. Usevich, "Structured low-rank approximation with missing data," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 2, pp. 814–830, 2013.

[27] Z. Liu, A. Hansson, and L. Vandenberghe, "Nuclear norm system identification with missing inputs and outputs," *Syst. Control Lett.*, vol. 62, no. 8, pp. 605–612, 2013.

[28] I. Markovsky, "Exact system identification with missing data," in *Proc. 52nd IEEE Conf. Decis. Control*, 2013, pp. 151–155.

[29] I. Markovsky, "The most powerful unfalsified model for data with missing values," *Syst. Control Lett.*, vol. 95, pp. 53–61, 2016.

[30] I. Markovsky and F. Dörfler, "Identifiability in the behavioral setting," *IEEE Trans. Autom. Control*, vol. 68, no. 3, pp. 1667–1677, Mar. 2022.

[31] I. Markovsky, E. Prieto-Araujo, and F. Dörfler, "On the persistency of excitation," *Automatica*, vol. 47, 2023, Art. no. 110657.

[32] L. N. Trefethen and D. Bau, *Numer. Linear Algebra*, vol. 181. Philadelphia, PA, USA: SIAM, 2022.

[33] M. Alsalti, M. Barkey, V. G. Lopez, and M. A. Müller, "Sample- and computationally efficient data-driven predictive control," in *Proc. 22nd IEEE Eur. Control Conf.*, 2020, pp. 307–313.

[34] I. Markovsky, M. Alsalti, V. G. Lopez, and M. A. Müller, "Identification from data with periodically missing output samples," *Automatica*, to be published.

[35] I. Krauss, V. G. Lopez, and M. A. Müller, "Sample-based observability of linear discrete-time systems," in *Proc. IEEE 61st Conf. Decis. Control*, 2022, pp. 4199–4205.

[36] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, no. 4, pp. 891–909, 2008.

[37] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[38] A. Padoan, J. Coulson, H. J. van Waarde, J. Lygeros, and F. Dörfler, "Behavioral uncertainty quantification for data-driven control," in *Proc. IEEE 61st Conf. Decis. Control*, 2022, pp. 4726–4731.

[39] A. Fazzi and I. Markovsky, "Distance problems in the behavioral setting," *Eur. J. Control*, vol. 74, 2023, Art. no. 100832.

[40] J. Miao and A. Ben-Israel, "On principal angles between subspaces in $\mathbb{R}^n$," *Linear Algebra Appl.*, vol. 171, pp. 81–98, 1992.

[41] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014.

[42] M. W. Causey, S. Miller, A. Foster, A. Beekley, D. Zenger, and M. Martin, "Validation of noninvasive hemoglobin measurements using the masimo Radical-7 SpHb station," *Amer. J. Surg.*, vol. 201, no. 5, pp. 592–598, 2011.

[43] C.-A. Ewaldsson and R. G. Hahn, "Kinetics and extravascular retention of acetated ringer's solution during isoflurane or propofol anesthesia for thyroid surgery," *J. Amer. Soc. Anesthesiologists*, vol. 103, no. 3, pp. 460–469, 2005.

[44] I. Markovsky, "A software package for system identification in the behavioral setting," *Control Eng. Pract.*, vol. 21, pp. 1422–1436, 2013.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.



Mohammad Alsalti (Graduate Student Member, IEEE) received the B.Sc. degree from the University of Jordan, Amman, Jordan, in 2017, and the M.Sc. degree from the University of Maryland, College Park, MD, USA, both in mechanical engineering. He is currently working toward the Ph.D. degree in electrical engineering with the Institute of Automatic Control, Leibniz University Hannover, Hanover, Germany.

In 2017, he was an Intern with NASA Ames Research Center as part of the intelligent robotics group. He is currently working on developing data-driven control techniques for linear and nonlinear systems with Leibniz University Hannover.

Mr. Alsalti was awarded the Fulbright scholarship to pursue his graduate studies at the University of Maryland (2018–2020).



Ivan Markovsky (Member, IEEE) received the Ph.D. degree in electrical engineering from the Katholieke Universiteit Leuven, Leuven, Belgium, in 2005.

He is currently an ICREA Professor with the International Centre for Numerical Methods in Engineering, Barcelona, Spain. From 2006 to 2012, he was an Assistant Professor with the School of Electronics and Computer Science, University of Southampton, Southampton, U.K., and from 2012 to 2022, an Associate Professor with the Vrije Universiteit, Brussel, Belgium. His research interests include computational methods for system theory, identification, and data-driven control in the behavioral setting.

Dr. Markovsky was the recipient of an ERC starting grant "Structured low-rank approximation: Theory, algorithms, and applications" 2010–2015, Householder Prize honorable mention 2008, and research mandate by the Vrije Universiteit Brussel research council 2012–2022.



Victor G. Lopez (Member, IEEE) received the B.Sc. degree in communications and electronics engineering from the Universidad Autonoma de Campeche, Campeche, Mexico, in 2010, the M.Sc. degree in electrical engineering from the Research and Advanced Studies Center (Cinvestav), Guadalajara, Mexico, in 2013, and the Ph.D. degree in electrical engineering from the University of Texas at Arlington (UTA), Texas, USA, in 2019.

Dr. Lopez was a Lecturer with the Western Technological Institute of Superior Studies (ITESO), Guadalajara, Mexico, in 2015. From 2019 to 2020, he was a Postdoctoral Researcher with the UTA Research Institute and an Adjunct Professor with the Electrical Engineering Department, UTA. He is currently a Postdoctoral Researcher with the Institute of Automatic Control, Leibniz University Hannover, Hannover, Germany. His research interest include cyber-physical systems, reinforcement learning, game theory, distributed control, and robust control.



Matthias A. Müller received the diploma degree in engineering cybernetics from the University of Stuttgart, Stuttgart, Germany, and the M.Sc. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, both in 2009, and the Ph.D. degree from the University of Stuttgart, in 2014.

Since 2019, he has been the Director of the Institute of Automatic Control and Full Professor with Leibniz University Hannover, Hanover, Germany. His research interests include nonlinear control and estimation, model predictive control, and data- and learning-based control, with applications in different fields including biomedical engineering and robotics.

Dr. Müller has received various awards for his work, including the 2015 European Systems & Control Ph.D. Thesis Award, the inaugural Brockett–Willems Outstanding Paper Award for the best paper published in Systems & Control Letters in the period 2014–2018, an ERC starting grant in 2020, the IEEE CSS George S. Axelby Outstanding Paper Award 2022, and the Journal of Process Control Paper Award 2023. He serves as an Associate Editor for *Automatica* and an Editor of the *International Journal of Robust and Nonlinear Control*.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:31:12 UTC from IEEE Xplore. Restrictions apply.