Identifiability in the Behavioral Setting

Ivan Markovsky and Florian Dörfler, *Senior Member, IEEE*

***Abstract*—Identifiability, i.e., uniqueness of a solution of the identification problem, is a fundamental issue in system identification and data-driven control. Necessary and sufficient identifiability conditions for deterministic linear time-invariant systems that do not require a priori given input/output partitioning of the variables nor controllability of the true system are derived in the article. The prior knowledge needed for identifiability is the number of inputs, lag, and order of the true system. Our results are based on a modification of the notion of a most powerful unfalsified model for finite data and a novel algorithm for its computation. We provide a generalization of a result that became known as the *fundamental lemma* and a novel nonparametric data-driven representation of the system behavior based on general data matrix structures. The results assume exact data, however, low-rank approximation allows their application in the case of noisy data. We compare empirically low-rank approximation of the Hankel, Page, and trajectory matrices in the errors-in-variables setting. Although the Page and trajectory matrices are unstructured, the parameter estimates obtained are less accurate than the one obtained from the Hankel matrix.**

***Index Terms*—Behavioral system theory, data-driven representations, most powerful unfalsified model, system identification.**

I. Introduction

THE *exact identification problem* aims to recover a dynamical system back from prior knowledge and data of the system. The *data* are one or more observed trajectories of the system. The prior knowledge is given as properties of the system, for example linearity, time-invariance, and bounded complexity. It is encoded in a *model class,* i.e., a set of systems to which the unknown data-generating system is a priori known to belong.

Manuscript received 27 July 2021; revised 28 July 2021 and 12 April 2022; accepted 20 April 2022. Date of publication 27 September 2022; date of current version 28 February 2023. This work was supported in part by the Catalan Institution for Research and Advanced Studies (ICREA), in part by the Fond for Scientific Research Vlaanderen (FWO) Projects G090117 N and G033822 N, and and in part the Fond Recherche Scientifique FNRS–FWO EOS Project 30468160. Recommended by Associate Editor M. Kanat Camlibel. *(Corresponding author: Ivan Markovsky.)*

Ivan Markovsky is with the Catalan Institution for Research and Advanced Studies (ICREA), 08010 Barcelona, Spain, and also with the International Centre for Numerical Methods in Engineering (CIMNE), Gran Capitàn, 08034 Barcelona, Spain (e-mail: imarkovsky@cimne.upc.edu).

Florian Dörfler is with the Automatic Control Laboratory (IfA), ETH-Zürich, 8092 Zürich, Switzerland (e-mail: dorfler@ethz.ch).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2022.3209954.

Digital Object Identifier 10.1109/TAC.2022.3209954

We consider discrete-time linear time-invariant deterministic systems and study conditions for well-posedness of the identification problem, i.e., conditions under which the problem has a unique solution. We refer to these conditions as *identifiability conditions*. In order to formally define the problem and derive its solution we use the behavioral approach to system theory.

In the behavioral approach, a dynamical system is defined as a set of trajectories—the *behavior* of the system. Properties of the system and problems involving the system are then defined in terms of the behavior. For example, the system is linear if the behavior is a subspace of the set of trajectories. The behavioral approach provides a general and rigorous language for dynamical phenomena. It was introduced in the early 1980s by Jan C. Willems as a result of resolving issues related to the classical view of a dynamical system as an input–output map.

There is an important difference between identifiability in the behavioral setting [1] and identifiability in the classical setting [2, Def. 4.6]. In the classical setting, the model is identified with a particular parametric representation and identifiability is defined as uniqueness of solution of a parameter estimation problem. The classical notion of identifiability depends on the *choice of the representation* as well as on the informativity of the data. In the behavioral setting, identifiability is defined in terms of the behavior, so that it is independent of the choice of the representation. Separating the issue of the nonuniqueness of the parameter in the model representations from the issue of informativity of the data is in our view an important advantage of the behavioral notion of identifiability.

A result, referred to as the *fundamental lemma* [1], provides the following identifiability conditions: 1) The data-generating system is controllable, and 2) an input component of the data is persistently exciting of a "sufficiently" high order. Implicit in the assumptions of the fundamental lemma is an a priori known input/output partitioning of the variables. Moreover, controllability is required, which is an overly restrictive assumption as it excludes for example the class of autonomous systems. Finally, the conditions are only sufficient. The fundamental lemma is originally derived for the case of a single given time series. It is generalized separately for multiple time series [3], the Page matrix [4], and uncontrollable systems [5]. The generalizations, however, still assume a given input/output partitioning, persistency of excitation of an input component, and provide only sufficient conditions.

Apart from giving identifiability conditions, the fundamental lemma has important practical applications. They stem from the fact that the fundamental lemma provides a *nonparametric data-driven representation* of a linear time-invariant system. Under the conditions of the fundamental lemma, the image

0018-9286 © 2022 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 08:34:11 UTC from IEEE Xplore. Restrictions apply.

of a Hankel matrix constructed from the data coincides with the set of all finite trajectories of the system. The data-driven representation led to subspace-type algorithms [6], [7]. See [8] for a comprehensive overview of data-driven analysis, signal processing, and control methods using the fundamental lemma.

An outstanding open problem for the application of the fundamental lemma in practice is dealing with approximation due to noisy data, disturbances, and model uncertainty. The result of [1] is for exact data and nontrivial modifications are needed for its application in case of noisy data. One approach for dealing with noise is using structured low-rank approximation [9], [10]. This leads to statistically optimal (ML solution) in the errors-in-variables setting [11], however, the resulting optimization problems are nonconvex. Subspace methods and convex relaxations based on the nuclear norm heuristic offer suboptimal solutions that can be used as initial approximation for the nonconvex optimization methods. Recently, new approaches for dealing with noise are developed in the context of data-driven control, see [8, Sec. 5.2].

The fundamental lemma can be viewed as an input design result—it provides guidelines for choosing the input that guarantee identifiability, independent of the initial conditions. Aiming to derive identifiability conditions that do not depend on a priori given input/output partition and controllability of the data-generating system led us to:

1) revision of the notion of the *most powerful unfalsified model* [12] for finite time series;
2) a novel algorithm for its computation;
3) new identifiability results;
4) a generalized data-driven representation.

These are the main contributions of this article. Although, as in the original work [1] on the fundamental lemma as well as in the follow-up publications [6], [7], [13], [14], [3], [5], [4], our theoretic results are derived assuming exact data, we relax and generalize the assumptions, demonstrating the

5) applicability of the theory for the case of noisy data.

We consider the errors-in-variables setting [11] and use as approximation methods low-rank approximation of the Hankel, Page, and trajectory matrices.

The rest of this article is organized as follows. Section II defines the notation and collects basic results used in the article. Although most of the material in this section is tutorial, an important new result is the formula ($\mathscr{B}|_L$, KER), which gives a matrix representation of the behavior restricted to a finite interval in terms of the parameters of a kernel representation. This result allows us to use linear algebra (instead of polynomial algebra) as the main tool for the analysis of linear time-invariant systems. Section III reviews the integer invariants of a linear time-invariant system. A core technical result of the article, on which the other results are based, is Lemma 4, which gives an explicit formula for the dimension of the behavior restricted to an interval of any length. In Section III-B, using Lemma 4 we define the notion of model complexity. Section IV introduces the notion of the most powerful unfalsified model. In Section IV-C, we present a modification of the most powerful unfalsified model for finite time series, which leads to a constructive exact system identification procedure (Algorithm 1). The main results of the article are in Section V. Theorem 17 gives necessary and sufficient identifiability conditions that do not require a priori known input/output partitioning and controllability of the true system. An application of the theory developed in the article, discussed in Section VI is using low-rank approximation of the unstructured Page and trajectory matrices for approximation in the presence of noise. Section VII shows simulation results of identification with noisy data in the errors-in-variables setting. Finally, Section VIII concludes this article.

II. Preliminaries

The goals of this section are to give the necessary background for the technical results in the rest of the article and to set the notation. We focus on the case of finite length, i.e., the restriction of the behavior to a finite interval. In Section II-A, we define the Hankel matrix as obtained by consecutive application of the shift and cut operators on the time series. Section II-B introduces the notions of a shift-invariant subspace and linear time-invariant system and defines the kernel, input/output, and input/state/output representations. Again, we focus on the finite length case. An important result of this section is the matrix characterization ($\mathscr{B}|_L$, KER) of the restricted behavior in terms of a kernel representation.

A. Hankel Matrices

We use interchangeably the terms *time series, sequence, and (discrete-time) trajectory*. The set of finite length q-variate, T-samples long, real-valued time series $w = (w(1), \ldots, w(T))$, where $w(t) \in \mathbb{R}^q$ is denoted by $(\mathbb{R}^q)^T$. The set of infinite length q-variate, real-valued time series $w = (w(1), w(2), \ldots)$, where $w(t) \in \mathbb{R}^q$ is denoted by $(\mathbb{R}^q)^{\mathbb{N}}$.

Selection of a part of a time series, defined over a subinterval of the time interval, is called *restriction*. The corresponding operator is called the *cut*. For a time series $w \in (\mathbb{R}^q)^T$ or $w \in (\mathbb{R}^q)^{\mathbb{N}}$ and an integer L, $1 \leq L \leq T$, we define the *cut operator*

$$w|_L := (w(1), \ldots, w(L)) \in (\mathbb{R}^q)^L. \qquad (\mathrm{w}|_\mathrm{L})$$

Applied on a set of time series $\mathscr{W} \subset (\mathbb{R}^q)^T$ or $\mathscr{W} \subset (\mathbb{R}^q)^{\mathbb{N}}$, the cut operator acts on all time series in the set, i.e.,

$$\mathscr{W}|_L := \{\, w|_L \mid w \in \mathscr{W} \,\} \subset (\mathbb{R}^q)^L.$$

With some abuse of notation we identify the finite sequence $(w|_L)$ with the vector $\mathrm{col}(w(1), \ldots, w_\mathrm{d}(L)) \in \mathbb{R}^{qL}$.

The derivative operator plays a key role for continuous-time dynamical systems. Its discrete-time equivalent is the *unit shift operator* $\sigma : (\mathbb{R}^q)^{\mathbb{N}} \mapsto (\mathbb{R}^q)^{\mathbb{N}}$ defined by $(\sigma w)(t) := w(t+1)$. Applied on a set of time series $\mathscr{W} \subset (\mathbb{R}^q)^{\mathbb{N}}$, σ acts on all elements of $\mathscr{W}$, i.e., $\sigma\mathscr{W} := \{\, \sigma w \mid w \in \mathscr{W} \,\}$. For a finite length time series $w \in (\mathbb{R}^q)^T$ and an integer $0 \leq \tau \leq T-1$, $\sigma^\tau w := (w(\tau+1), \ldots, w(T))$, so that σ can be applied to $w \in (\mathbb{R}^q)^T$ at most $T-1$ times. The resulting time series put next to each other form the 2-D array of numbers shown in Table I.Question marks indicate unspecified elements.

Restriction of $w, \sigma w, \ldots, \sigma^{T-L} w$ to the interval $[1, L]$, for some $1 \leq L \leq T$, is equivalent to consecutive application of

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:34:11 UTC from IEEE Xplore. Restrictions apply.

TABLE I
CONSECUTIVE APPLICATION OF THE SHIFT OPERATOR ON A FINITE TIME SERIES RESULTS IN A HANKEL MATRIX WITH MISSING VALUES

w	σw	$\cdots$	$\sigma^{T-1}w$
$w(1)$	$w(2)$	$\cdots$	$w(T)$
$w(2)$	$\vdots$	$\iddots$	?
$\vdots$	$w(T)$	$\iddots$	$\vdots$
$w(T)$	?	$\cdots$	?

the shift and cut operators. The *shift-and-cut* operator is used in [15] for state construction. Here we use it for defining the *Hankel matrix* of depth L

$$\mathscr{H}_L(w) := \begin{bmatrix} w|_L & (\sigma w)|_L & \cdots & (\sigma^{T-L}w)|_L \end{bmatrix} \in \mathbb{R}^{qL\times(T-L+1)}$$

which is a fully defined upper-left block in Table I.

Although $\mathscr{H}_L(w)$ is defined for all integers $L \in [1, T]$, we will require that $\mathscr{H}_L(w)$ has at least as many columns as rows. This limits the range of values for L to $[1, L_{\max}]$, where

$$L_{\max} := \left\lfloor \tfrac{T+1}{q+1} \right\rfloor \qquad (\mathrm{L_{max}})$$

is the largest integer smaller than or equal to $(T+1)/(q+1)$.

In the case of an infinite sequence w, $\mathscr{H}_L(w)$ is the one-side infinite Hankel matrix

$$\mathscr{H}_L(w) = \begin{bmatrix} w|_L & (\sigma w)|_L & (\sigma^2 w)|_L & \cdots \end{bmatrix} \in \mathbb{R}^{qL\times\infty}.$$

We define also the two-side infinite Hankel matrix

$$\mathscr{H}(w) = \begin{bmatrix} w & \sigma w & \sigma^2 w & \cdots \end{bmatrix} \in \mathbb{R}^{\infty\times\infty}.$$

The *Page matrix* [16] $\mathscr{P}_L(w)$ of w with L block rows is obtained from the Hankel matrix $\mathscr{H}_L(w)$ by column selection

$$\mathscr{P}_L(w) := \begin{bmatrix} w|_L & (\sigma_L w)|_L & \cdots & (\sigma_{(T'-1)L} w)|_L \end{bmatrix} \in \mathbb{R}^{qL\times T'}$$

$$= \begin{bmatrix} w_1 & w_{L+1} & \cdots & w_{(T'-1)L+1} \\ \vdots & \vdots & & \vdots \\ w_L & w_{2\,L} & \cdots & w_{T'L} \end{bmatrix} \qquad (\mathscr{P}_{\mathrm{L}}(\mathrm{w}))$$

where $T' := \lfloor T/L \rfloor$. Like the Hankel matrix $\mathscr{H}_L(w)$, the Page matrix $\mathscr{P}_L(w)$ also consists of L-samples long trajectories, however, unlike the Hankel matrix, the Page matrix has no repeated elements on the antidiagonals.

The operator $\mathscr{P}_L$ constructing the Page matrix is called the *lifting operator*. It is used for identification of linear periodically time-varying systems [17] as well as an alternative to the Hankel matrix $\mathscr{H}_L(w)$ for data-driven control [4], approximate realization of linear time-invariant systems [16], and time series analysis [18].

A generalization of the Hankel matrix for a set

$$\mathscr{W} := \{ w^1, \ldots, w^N \}, \quad w^i \in (\mathbb{R}^q)^{T_i}$$

of N time series is the *mosaic-Hankel matrix* [19], [20], [3]

$$\mathscr{H}_L(\mathscr{W}) := \begin{bmatrix} \mathscr{H}_L(w^1) & \cdots & \mathscr{H}_L(w^N) \end{bmatrix} \in \mathbb{R}^{qL\times\sum_{i=1}^N (T_i-L)}. \qquad (\mathscr{H}_{\mathrm{L}}(\mathscr{W}))$$

The range of values for L is $[1, L_{\max}]$, where $L_{\max}$ is defined as in $(L_{\max})$ with $T := \max\{ T_1, \ldots, T_N \}$. Blocks $\mathscr{H}_L(w^i)$ in $(\mathscr{H}_L(\mathscr{W}))$, for which $T_i < L$ {are not well defined, so that they are dropped.}

A special case of the mosaic-Hankel matrix when all time series w^i have length $T_1 = \cdots = T_N = L$, i.e, $w^i \in (\mathbb{R}^q)^L$ for all $i \in \{ 1, \ldots, N \}$, is the *trajectory matrix*

$$\mathscr{T}(\mathscr{W}) := \begin{bmatrix} w_1^1 & w_1^2 & \cdots & w_1^N \\ \vdots & \vdots & & \vdots \\ w_L^1 & w_L^2 & \cdots & w_L^N \end{bmatrix} \in \mathbb{R}^{qL\times N}. \qquad (\mathscr{T}(\mathscr{W}))$$

The trajectory matrix is used for dictionary learning [21]. The Page matrix $\mathscr{P}_L(w)$ is a special trajectory matrix obtained by taking $w^i = (\sigma^{(i-1)L} w)|_L$, for $i \in \{ 1, \ldots, T' \}$.

B. Shift-Invariant Subspaces and Linear Time-Invariant Systems

In the behavioral setting, *dynamical systems* are sets of trajectories. The trajectories of a discrete-time system are time series, so that a discrete-time system $\mathscr{B}$ is a subset of $(\mathbb{R}^q)^{\mathbb{N}}$. The system $\mathscr{B} \subset (\mathbb{R}^q)^{\mathbb{N}}$ is linear if $\mathscr{B}$ is a subspace and time-invariant if $\mathscr{B}$ is shift-invariant.

Definition 1 (Shift-invariant spaces): The set $\mathscr{W} \subset (\mathbb{R}^q)^{\mathbb{N}}$ is *shift-invariant* if for any $w \in \mathscr{W}$ and $\tau \in \mathbb{N}$, $\sigma^\tau w \in \mathscr{W}$. The set $\mathscr{W} \subset (\mathbb{R}^q)^T$ is shift-invariant if for any $w \in \mathscr{W}$ and $\tau \in \{ 0, 1, \ldots, T-1 \}$, there is $v \in \mathscr{W}$, such that $\sigma^\tau w = v|_{T-\tau}$.

A classic example of a shift-invariant space is the image of the observability matrix $\mathscr{O}_T(A, C) := \mathrm{col}(C, CA, \ldots, CA^{T-1})$. The set of linear time-invariant systems with q variables is denoted by $\mathscr{L}^q$. A finite-dimensional linear time-invariant system $\mathscr{B}$ admits a *kernel representation*

$$\mathscr{B} = \ker\, R(\sigma) := \{ w \mid R(\sigma) w = 0 \} \qquad (\mathrm{KER})$$

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

$$R(z) = R_0 + R_1 z + \cdots + R_\ell z^\ell$$

$$= \begin{bmatrix} R^1(z) \\ \vdots \\ R^g(z) \end{bmatrix} = \begin{bmatrix} R_0^1 + R_1^1 z + \cdots + R_{\ell_1}^1 z^{\ell_1} \\ \vdots \\ R_0^g + R_1^g z + \cdots + R_{\ell_g}^g z^{\ell_g} \end{bmatrix} \in \mathbb{R}^{g\times q}[z]. \qquad (\mathrm{R})$$

(KER) is called *minimal* if the number of equations g is as small as possible over all kernel representations of $\mathscr{B}$. As shown in Proposition 3, in a minimal kernel representation $g = p$, i.e., the number of equations is equal to the number of outputs. At the same time, in a minimal kernel representation, $\deg R$ is minimized over all kernel representations of $\mathscr{B}$, and the minimal $\deg R$ is equal to the lag of the system $\mathscr{B}$.

Definition 2 (Annihilator): An operator $r(\sigma)$, $r(z) \in \mathbb{R}^{1\times q}[z]$, such that $r(\sigma)\mathscr{B} = 0$ is called an *annihilator* of $\mathscr{B}$.

For example, $r(\sigma) = 1 + \sigma - \sigma^2$ is an annihilator of the Fibonacci sequence $(1, 1, 2, 3, 5, \ldots)$. In general, the rows $R^1, \ldots, R^g$ of R are annihilators of $\mathscr{B} = \ker\, R(\sigma)$. Moreover, the rows of R form a basis for all annihilators of $\mathscr{B}$. The *multiplication matrix* of $r = r_0 + r_1 z + \cdots + r_\ell z^\ell \in \mathbb{R}^{1\times q}[z]$

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:34:11 UTC from IEEE Xplore. Restrictions apply.

with $L > \ell := \deg\, r$ block columns is the matrix

$$\mathscr{M}_L(r) := \begin{bmatrix} r_0 & r_1 & \cdots & r_\ell & & & \\ & r_0 & r_1 & \cdots & r_\ell & & \\ & & \ddots & \ddots & & \ddots & \\ & & & r_0 & r_1 & \cdots & r_\ell \end{bmatrix} \in \mathbb{R}^{(L-\ell)\times qL}.$$

For $L \leq \ell$, we define $\mathscr{M}_L(r)$ to be a $0 \times qL$ empty matrix, $\operatorname{rank} \mathscr{M}_L(r) = 0$, and $\ker\, \mathscr{M}_L(r) = \mathbb{R}^{qL}$.

Consider the polynomial operator $R(\sigma)$ in a kernel representation (KER). We have the following matrix theoretic characterization of the restricted behavior of the system $\mathscr{B} = \ker\, R(\sigma)$ to the interval $[1, L]$ in terms of (KER)

$$\mathscr{B}|_L = \ker \begin{bmatrix} \mathscr{M}_L(R^1) \\ \vdots \\ \mathscr{M}_L(R^g) \end{bmatrix} =: \ker\, \mathscr{M}_L(R). \qquad (\mathscr{B}|_{\mathrm{L}}, \mathrm{KER})$$

For a permutation matrix $\Pi \in \mathbb{R}^{q\times q}$ and an integer $0 < m < q$

$$\begin{bmatrix} u \\ y \end{bmatrix} := \Pi^{-1} w \qquad (\mathrm{w} \mapsto (\mathrm{u}, \mathrm{y}))$$

define a *partitioning* of the variables $w(t) \in \mathbb{R}^q$ into $u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}^{q-m}$. Let Π_u be the projection of w on the u variable, i.e., $\Pi_u w := u$. Acting on a set, Π_u projects all elements in the set, which results in a new set. The partitioning $(w \mapsto (u, y))$ is an *input/output partitioning* of $\mathscr{B}$ if 1. u is free, i.e., $\Pi_u \mathscr{B} = (\mathbb{R}^m)^{\mathbb{N}}$, 2. $\dim\, u$ is maximal, and 3. y does not anticipate u, i.e., $w^1, w^2 \in \mathscr{B}$ and $u^1(t) = u^2(t)$ for $t \leq 0$ imply that there is y^2, such that $\Pi \begin{bmatrix} u^2 \\ y^2 \end{bmatrix} \in \mathscr{B}$ and $y^2(t) = y^1(t)$, $t \leq 0$.

Let $\mathscr{B} = \ker\, R(\sigma)$ be a minimal kernel representation of $\mathscr{B}$. The partitioning $(w \mapsto (u, y))$ is an *input/output partitioning* of $\mathscr{B}$ if and only if with $\begin{bmatrix} Q & -P \end{bmatrix} := R\Pi^{-1}$, with $P \in \mathbb{R}^{p\times p}$ nonsingular [22]. The resulting *input/outut representation* is

$$\mathscr{B}_{\mathrm{i/o}}(Q, P, \Pi) = \left\{ \Pi \begin{bmatrix} u \\ y \end{bmatrix} \mid Q(\sigma)u = P(\sigma)y \right\}. \qquad (\mathrm{I/O})$$

A finite dimensional linear time-invariant system $\mathscr{B}$ admits an *input/state/output representation*

$$\mathscr{B} = \mathscr{B}_{\mathrm{ss}}(A, B, C, D, \Pi) := \left\{ \Pi \begin{bmatrix} u \\ y \end{bmatrix} \mid \text{there is } x \in (\mathbb{R}^n)^{\mathbb{N}}, \right.$$
$$\left. \text{such that } \sigma x = Ax + Bu,\ y = Cx + Du \right\}, \quad (\mathrm{I/S/O})$$

where $\Pi \in \mathbb{R}^{q\times q}$ is a permutation and $\begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathbb{R}^{(n+p)\times(n+m)}$. (I/S/O) is called *minimal* if $n := \dim\, A$ is as small as possible over all input/state/output representations of $\mathscr{B}$. Associated to $\mathscr{B} = \mathscr{B}_{\mathrm{ss}}(A, B, C, D, \Pi)$, is the *autonomous subbehavior*

$$\mathscr{B}_{\mathrm{ss}}(A, C) := \Big\{ y \mid \text{there is } x \in (\mathbb{R}^n)^{\mathbb{N}}, \text{ such that } \sigma x = Ax,\ y = Cx \Big\}.$$

The next section shows how the parameters n and ℓ appearing in minimal input/state/output and kernel representations are connected to the behavior.

III. Integer Invariants and Model Complexity

The concept of integer invariants [23] is a key one in the theory of linear time-invariant systems. The integer invariants are used in Lemma 4 to characterize the dimension of the restricted behavior over an interval of any length. A related result, stated as Corollary 5, gives the dimension of the restriction of the behavior over intervals of length larger than the lag. Based on Corollary 5, in Section III-B we define the model complexity as the triple: (number of inputs, lag, order).

A. Integer Invariants

For a linear time-invariant system $\mathscr{B}$ we define the following integers, called *invariants* of $\mathscr{B}$.

1) *Input cardinality* $\boldsymbol{m}(\mathscr{B})$:= number of inputs in (I/O).
2) *Output cardinality* $\boldsymbol{p}(\mathscr{B})$:= number of outputs in (I/O).
3) *Order* $\boldsymbol{n}(\mathscr{B})$:= minimal number of states in (I/S/O).
4) *Lag* $\boldsymbol{\ell}(\mathscr{B})$:= minimal degree of R in (KER), and
5) *Structure indices* $(\boldsymbol{\ell}_1(\mathscr{B}), \ldots, \boldsymbol{\ell}_p(\mathscr{B}))$, $\boldsymbol{\ell}_i(\mathscr{B}) := \deg\, R^i$ in a minimal kernel representation (KER) of $\mathscr{B}$. Without loss of generality, we assume $\boldsymbol{\ell}_1(\mathscr{B}) \leq \cdots \leq \boldsymbol{\ell}_p(\mathscr{B})$ and define $\boldsymbol{\ell}_0(\mathscr{B}) := 0$, $\boldsymbol{\ell}_{p+1}(\mathscr{B}) := \infty$ for any $\mathscr{B}$.

In order to be well defined, the invariants of $\mathscr{B}$ should not depend on the choice of the representation.

Proposition 3 (Wellposedness, Proposition X.3 [22]): Let $\mathscr{B} = \ker\, R(\sigma)$ be a minimal kernel representation and $\mathscr{B} = \mathscr{B}_{\mathrm{ss}}(A, B, C, D, \Pi)$ be a minimal input/state/output representation of a linear time-invariant system $\mathscr{B}$. Then, $\boldsymbol{m}(\mathscr{B})$, $\boldsymbol{p}(\mathscr{B})$, $\boldsymbol{n}(\mathscr{B})$, $\boldsymbol{\ell}(\mathscr{B})$, and $(\boldsymbol{\ell}_1(\mathscr{B}), \ldots, \boldsymbol{\ell}_p(\mathscr{B}))$ are invariant of the representations. Moreover,

1) $\boldsymbol{p}(\mathscr{B}) = \operatorname{row\,dim}\, C = \operatorname{row\,dim}\, R$;
2) $\boldsymbol{m}(\mathscr{B}) = \operatorname{col\,dim}\, B = q - \operatorname{row\,dim}\, R$;
3) $\boldsymbol{\ell}(\mathscr{B}) = \max\{\, \boldsymbol{\ell}_1(\mathscr{B}), \ldots, \boldsymbol{\ell}_p(\mathscr{B}) \,\}$; and
4) $\boldsymbol{n}(\mathscr{B}) = \boldsymbol{\ell}_1(\mathscr{B}) + \cdots + \boldsymbol{\ell}_p(\mathscr{B})$.

Note that $(\boldsymbol{\ell}_0(\mathscr{B}), \boldsymbol{\ell}_1(\mathscr{B})] \cup \cdots \cup (\boldsymbol{\ell}_p(\mathscr{B}), \boldsymbol{\ell}_{p+1}(\mathscr{B}))$ define a partitioning of $\mathbb{N}$, so that for any $L \in \mathbb{N}$, there is $k(L) \in \{\, 0, 1, \ldots, p \,\}$, such that $L \in (\boldsymbol{\ell}_{k(L)}(\mathscr{B}), \boldsymbol{\ell}_{k(L)+1}(\mathscr{B})]$.

The following core technical lemma shows that $\mathscr{B}|_L$ is a piecewise affine function of L with kinks at the structure indices. Initially, $\mathscr{B}|_L$ increases at the rate of q. At each kink point the rate of increase drops by 1. For $L > \ell$, the rate reaches the constant $\boldsymbol{m}(\mathscr{B}) = q - \boldsymbol{p}(\mathscr{B})$, see Fig. 1. □

Lemma 4 (Dimension of $\mathscr{B}|_L$): Let $\mathscr{B}$ be a linear time-invariant system. Then,

$$\dim \mathscr{B}|_L = (q - k(L))\, L + \boldsymbol{\ell}_1(\mathscr{B}) + \cdots + \boldsymbol{\ell}_{k(L)}(\mathscr{B}). \quad (\dim\, \mathscr{B}|_{\mathrm{L}})$$

Proof: Consider a minimal kernel representation $\mathscr{B} = \ker\, R(\sigma)$. From $(\mathscr{B}|_L, \mathrm{KER})$, we have that

$$\dim\, \mathscr{B}|_L = qL - \operatorname{rank} \mathscr{M}_L(R).$$

Since $\mathscr{M}_L(R)$ is full row rank, we can find its rank from the row dimension $\operatorname{rank} \mathscr{M}_L(R) = \sum_{i=1}^{k(L)} (L - \ell_i) = k(L)L - \sum_{i=1}^{k(L)} \ell_i$. For $L < \ell_1$, $k(L) = 0$, $\mathscr{M}_L(R)$ is an empty matrix, and $\operatorname{rank} \mathscr{M}_L(R) = 0$, so that $\sum_{i=1}^{0} \ell_i = 0$. □

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:34:11 UTC from IEEE Xplore. Restrictions apply.

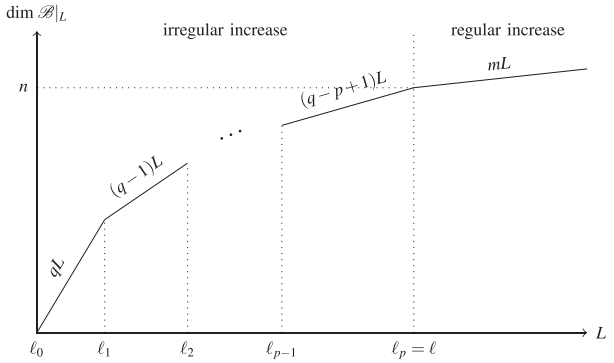


Fig. 1. Dimension of $\mathscr{B}|_L$ is a piecewise affine function of L with kinks at the structure indices $\ell_1, \ldots, \ell_p$.

Lemma 4 shows that after an irregular increase of the dimension of $\mathscr{B}|_L$ in the interval $[1, \ell]$, in the final stage $L \in [\ell, \infty)$, $\mathscr{B}|_L$ increases linearly at a rate $\boldsymbol{m}(\mathscr{B})$.

Indeed, $k(L) = \boldsymbol{p}(\mathscr{B})$ for $L \geq \ell$ and $\sum_{i=1}^{\boldsymbol{p}(\mathscr{B})} \ell_i = \boldsymbol{n}(\mathscr{B})$.

Corollary 5: For $L \geq \boldsymbol{\ell}(\mathscr{B})$, $\dim \mathscr{B}|_L$ is an affine function of L with slope determined by $\boldsymbol{m}(\mathscr{B})$ and offset $\boldsymbol{n}(\mathscr{B})$,

$$\dim \mathscr{B}|_L = \boldsymbol{m}(\mathscr{B})L + \boldsymbol{n}(\mathscr{B}), \text{ for all } L \geq \boldsymbol{\ell}(\mathscr{B}). \quad (\dim \mathscr{B}|'_{\mathrm{L}})$$

In the special case of an autonomous system, after an increase of the dimension of $\mathscr{B}|_L$ in the interval $[1, \ell]$, $\mathscr{B}|_L$ becomes constant and is equal to the order of $\mathscr{B}$. The fact that $\mathscr{B}|_L = \boldsymbol{n}(\mathscr{B})$, for an linear time-invariant autonomous system and $L \geq \boldsymbol{\ell}(\mathscr{B})$ is well known from realization theory.

Note 6: For $L = 1, \ldots, \ell$, $(\dim \mathscr{B}|_L)$ is a system of linear equations. Due to its staircase structure, it has a unique solution for $\ell_1, \ldots, \ell_p$ in terms of $\dim \mathscr{B}|_1, \ldots, \dim \mathscr{B}|_\ell$. This shows that the integer invariants are uniquely defined by $\mathscr{B}$.

B. Model Complexity

Linear systems are subspaces. The system's complexity is related to its dimension—the higher dimensional the subspace, the more complex the corresponding system. For a linear time-invariant system $\mathscr{B}$ with $\boldsymbol{m}(\mathscr{B}) > 0$ inputs, however, $\dim \mathscr{B} = \infty$, despite the fact that $\mathscr{B}$ admits representations with finite dimensional parameter vectors. Therefore, we characterize the complexity of $\mathscr{B}$ by the increase of the dimension of $\mathscr{B}|_L$ as L grows. As shown in Lemma 4, this increase is fully characterized by the structure indices $(\boldsymbol{\ell}_1(\mathscr{B}), \ldots, \boldsymbol{\ell}_p(\mathscr{B}))$. By Corollary 5, however, for $L > \boldsymbol{\ell}(\mathscr{B})$, the increase of $\mathscr{B}|_L$ depends only on $\boldsymbol{m}(\mathscr{B})$ and $\boldsymbol{n}(\mathscr{B})$.

Definition 7 (Model complexity): Complexity of a linear time-invariant system $\mathscr{B}$ is a triplet $\boldsymbol{c}(\mathscr{B}) := (\boldsymbol{m}(\mathscr{B}), \boldsymbol{\ell}(\mathscr{B}), \boldsymbol{n}(\mathscr{B}))$. Complexities are compared by lexicographic ordering: a model with more inputs is more complex than a model with fewer inputs irrespective of their lags and orders.

The class of bounded complexity linear time-invariant systems is defined as $\mathscr{L}^{q,n}_{m,\ell} := \{ \mathscr{B} \in \mathscr{L}^q \mid \boldsymbol{c}(\mathscr{B}) \leq (m, \ell, n) \}$. In addition to systems with order n, lag ℓ, and m inputs, $\mathscr{L}^{q,n}_{m,\ell}$ includes all *lower complexity* systems—systems with fewer than m inputs, and/or lag smaller than ℓ, and/or order smaller than n. In order to fix the complexity of a system we use the notation $\partial\mathscr{L}^{q,n}_{m,\ell} := \{ \mathscr{B} \in \mathscr{L}^q \mid \boldsymbol{c}(\mathscr{B}) = (m, \ell, n) \}$.

Note 8 (Missing letters in $\mathscr{L}^{q,n}_{m,\ell}/\partial\mathscr{L}^{q,n}_{m,\ell}$): Missing letters n, ℓ, or m in the notation $\mathscr{L}^{q,n}_{m,\ell}$ mean that there are no bounds imposed on the corresponding system's invariants—order, lag, and input cardinality, respectively. "$\bullet$" can be used as a place holder for unrestricted quantities. For example, $\mathscr{L}^{q}_{m,\ell}$ is the set of linear time-invariant systems with at most m inputs and lag upper bounded by ℓ, while $\mathscr{L}^{q}_{\bullet,\ell}$ is the set of systems with an upper bound on the lag but no bound on the number of inputs.

The order $\boldsymbol{n}(\mathscr{B})$ of a system $\mathscr{B} \in \mathscr{L}^q$ must satisfy the implicit constraints $\boldsymbol{\ell}(\mathscr{B}) \leq \boldsymbol{n}(\mathscr{B}) \leq \boldsymbol{p}(\mathscr{B})\boldsymbol{\ell}(\mathscr{B})$. Therefore, $\mathscr{L}^{q,n}_{m,\ell} \subseteq \mathscr{L}^{q}_{m,\ell}$, i.e., the model class $\mathscr{L}^{q,n}_{m,\ell}$ specifies a more refined structure than $\mathscr{L}^{q}_{m,\ell}$.

Similarly, missing letters in the notation $\partial\mathscr{L}^{q,n}_{m,\ell}$ imply that the corresponding invariants are not restricted.

IV. Most Powerful Unfalsified Model

The "clear and rational foundation under the problem of obtaining models from time series," developed by Willems in [23], culminates in the notion of the *most powerful unfalsified model*. The definition of the most powerful unfalsified model $\mathscr{B}_{\mathrm{MPUM}}(\mathscr{W}_{\mathrm{d}})$ for the data $\mathscr{W}_{\mathrm{d}}$ in the model class $\mathscr{L}^q$, recalled in Section IV-A, is based on the notion of model complexity defined in Section III-B. The complexity (m, ℓ, n) is minimized in the lexicographic order over the constraint that the model is exact.

In [23], infinite length data are considered. In the case of finite length data, in Section IV-B we show that independent of the data $\mathscr{W}_{\mathrm{d}}$, $\mathscr{B}_{\mathrm{MPUM}}(\mathscr{W}_{\mathrm{d}})$ is a trivial autonomous system (Lemma 11). This critical issue is overlooked in the literature. In order to avoid the trivial solution a bound on the model complexity should be imposed.

Section IV-C presents a modification of the most powerful unfalsified model for finite data and an algorithm for computing a minimal kernel representation for it. As a side result, we show that if two systems have behaviors that coincide over an interval of length L, where L is larger than the lags of the systems, then the systems coincide.

A. Infinite Length Case

Definition 9: The most powerful unfalsified model $\mathscr{B} = \mathscr{B}_{\mathrm{MPUM},\mathscr{M}}(\mathscr{W}_{\mathrm{d}})$ for the data $\mathscr{W}_{\mathrm{d}} \subset (\mathbb{R}^q)^{\mathbb{N}}$ in the model class $\mathscr{M}$ is defined by the following properties.

1) $\mathscr{B}$ is *in the model class,* i.e., $\mathscr{B} \in \mathscr{M}$.
2) $\mathscr{B}$ is *unfalsified,* i.e., $\mathscr{W}_{\mathrm{d}} \subseteq \mathscr{B}$.
3) $\mathscr{B}$ is *most powerful,* i.e., any other model $\mathscr{B}'$ for which properties 1 and 2 hold is no more powerful than $\mathscr{B}$, i.e.,

$$\boldsymbol{c}(\mathscr{B}) \leq \boldsymbol{c}(\mathscr{B}'). \qquad (\mathscr{B} \subseteq \mathscr{B}')$$

Definition 9 leads to the multi-objective optimization problem $\min_{\mathscr{W}_{\mathrm{d}} \subseteq \mathscr{B} \in \mathscr{M}} \boldsymbol{c}(\mathscr{B})$. In this article, the model classes considered are $\mathscr{L}^q$, $\mathscr{L}^{q,n}_{m,\ell}$, and $\partial\mathscr{L}^{q,n}_{m,\ell}$. The complexity can be upper bounded or fixed by the *hyperparameters* m, ℓ, and n. If the model class is the unrestricted $\mathscr{L}^q$ (i.e., hyperparameters are

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:34:11 UTC from IEEE Xplore. Restrictions apply.

not specified), we use the notation $\mathscr{B}_{\text{MPUM}}$. The dataset $\mathscr{W}_{\text{d}}$ may consist of a single trajectory $w_{\text{d}} \in (\mathbb{R}^q)^{\mathbb{N}}$ or of multiple trajectories, $\mathscr{W}_{\text{d}} = \{ w_{\text{d}}^1, \ldots, w_{\text{d}}^N \}$, $w_{\text{d}}^i \in (\mathbb{R}^q)^{\mathbb{N}}$.

The most powerful unfalsified model in $\mathscr{L}^q$ always exists and is unique. Indeed, $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}}) = \text{image}\, \{ \mathscr{W}_{\text{d}} \cup \sigma\mathscr{W}_{\text{d}} \cup \sigma^2\mathscr{W}_{\text{d}} \cup \cdots \}$. An insightful view of this relation is as the image of the Hankel matrix constructed from the data. For a dataset consisting of a single trajectory $\mathscr{W}_{\text{d}} = \{ w_{\text{d}} \}$, $w_{\text{d}} \in (\mathbb{R}^q)^{\mathbb{N}}$

$$\mathscr{B}_{\text{MPUM}}(w_{\text{d}}) = \text{image}\, \mathscr{H}(w_{\text{d}}).$$

For multiple trajectories $\mathscr{W}_{\text{d}} = \{ w_{\text{d}}^1, \ldots, w_{\text{d}}^N \}$, $\mathscr{B}_{\text{MPUM}}(w_{\text{d}})$ is given by the image of the mosaic-Hankel matrix

$$\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}}) = \text{image}\, \underbrace{\begin{bmatrix} \mathscr{H}(w_{\text{d}}^1) & \cdots & \mathscr{H}(w_{\text{d}}^N) \end{bmatrix}}_{\mathscr{H}(\mathscr{W}_{\text{d}})}.$$

All subsequent results in the article (for finite as well as infinite data) hold true for multiple trajectories by using mosaic-Hankel matrices in place of Hankel matrices.

Although, by definition $\mathscr{B}_{\text{MPUM}}$ imposes a priori no bound on the complexity, depending on the data $\mathscr{W}_{\text{d}}$, the actual model $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}})$ may have bounded complexity.

Lemma 10 (Complexity of $\mathscr{B}_{MPUM}(\mathscr{W}_d)$): Given a set of time series $\mathscr{W}_{\text{d}} \subset (\mathbb{R}^q)^{\mathbb{N}}$, let ℓ be the smallest natural number, such that there are m and n satisfying the equations

$$\dim \text{ image } \mathscr{H}_L(\mathscr{W}_{\text{d}}) = mL + n, \text{ for } L \in \{ \ell, \ell+1, \ldots \}. \quad (\dim)$$

Then, $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}}) \in \partial\mathscr{L}^{q,n}_{m,\ell}$.

Proof: By definition, $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}}) = \text{image}\, \mathscr{H}(\mathscr{W}_{\text{d}})$. Since

$$\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}})|_L = \text{image}\, \mathscr{H}_L(\mathscr{W}_{\text{d}})$$

by (dim) and ($\dim\, \mathscr{B}|_L$), we have $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}}) \in \partial\mathscr{L}^{q,n}_{m,\ell}$. □

Contrary to $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}})$, which always exists, the most powerful unfalsified model $\mathscr{B}_{\text{MPUM},\mathscr{L}^{q,n}_{m,\ell}}(\mathscr{W}_{\text{d}})$ in a model class of bounded complexity $\mathscr{L}^{q,n}_{m,\ell}$ may not exist.

B. Finite Length Case

In the finite length case without complexity specification, Definition 9 leads to an autonomous model.

Lemma 11: For finite length data, $\mathscr{B}_{MUPU}(\mathscr{W}_{\text{d}})$ is autonomous, i.e., $\boldsymbol{m}(\mathscr{B}_{MPUM}(\mathscr{W}_{\text{d}})) = 0$.

Proof: By the lexicographic ordering of the complexity, any autonomous system is less complex than any system with inputs. Therefore, for the proof it suffices to show that there exists an exact autonomous model. Consider first a single finite trajectory $w_{\text{d}} \in (\mathbb{R}^q)^T$. The claim is that there is an autonomous linear time-invariant system $\mathscr{B}$ that fits w_{d} exactly. Let $\mathscr{B} = \mathscr{B}_{\text{ss}}(A, C)$ be a minimal state space representation of the model. Then, $w_{\text{d}} \in \mathscr{B}|_T$ if and only if the system of equations $w_{\text{d}} = \mathscr{O}x_{\text{ini}}$, where $\mathscr{O}$ is the extended observability matrix $\mathscr{O}_T(A, C)$, has a solution $x_{\text{ini}} \in \mathbb{R}^{\boldsymbol{n}(\mathscr{B})}$. There is a solution x_{ini}, for *any* $w_{\text{d}} \in (\mathbb{R}^q)^T$ if $\mathscr{O}_T(A, C)$ is full row rank. This can be guaranteed by choosing the order n sufficiently large. For example, take $n = qT$ and

$$C = \begin{bmatrix} I_q & 0 & \cdots & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & I_q & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_q \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Then, $\mathscr{O}_T(A, C) = I_{qT}$ and the solution is $w_{\text{d}} = x_{\text{ini}}$.

In case of a finite number of finite length trajectories $\{ w_{\text{d}}^1, \ldots, w_{\text{d}}^N \}$, the argument above can be applied repeatedly for each trajectory w_{d}^i, resulting in an exact autonomous model $\mathscr{B}^i = \mathscr{B}_{\text{ss}}(A^i, C^i)$. Define then $\mathscr{B}$ as $\mathscr{B} = \mathscr{B}_{\text{ss}}\left(\text{diag}(A^1, \ldots, A^N), \begin{bmatrix} C^1 & \cdots & C^N \end{bmatrix}\right)$, which is an exact autonomous model for $\mathscr{W}_{\text{d}}$. □

Lemma 11 shows a degeneracy: without a suitable bound on the complexity, in the finite length case, the most powerful unfalsified model always leads to a trivial model. This fact is observed in [24], where as a possible solution it is proposed to fix the number of inputs, i.e., (in our notation) the model $\mathscr{B}_{\text{MPUM},\partial\mathscr{L}^q_m}(\mathscr{W}_{\text{d}})$ is considered. Next, we modify the definition of the most powerful unfalsified model for the finite data case in order to avoid the trivial model when prior knowledge about the number of inputs, lag, or order is not given.

C. Modification and Algorithm

From the results in Section IV-B it seems that an upper bound on the complexity of the model should be a priori given in order to avoid the trivial model in the finite time case. In fact, prior knowledge about the complexity of the model need not be given by the hyperparameters (m, ℓ, n) because there is a universal upper bound on the lag in case of finite data

$$\mathscr{W}_{\text{d}} = \{ w_{\text{d}}^1, \ldots, w_{\text{d}}^N \}, \quad w_{\text{d}}^i \in (\mathbb{R}^q)^{T_i}. \qquad (\mathscr{W}_{\text{d}})$$

In order to see this, let $T := \max\{ T_1, \ldots, T_N \}$, define

$$L_{\max} := \left\lfloor \tfrac{T+1}{q+1} \right\rfloor$$

and recall Definition 2 of an annihilator. Based on the finite data $\mathscr{W}_{\text{d}}$, only annihilators of degree up to $L_{\max} - 1$ can be determined from $\mathscr{W}_{\text{d}}$. Indeed, an annihilator of degree $L - 1$ is computed from the left kernel of the Hankel matrix $\mathscr{H}_L(\mathscr{W}_{\text{d}})$, however, for $L > L_{\max}$ the left kernel of $\mathscr{H}_L(\mathscr{W}_{\text{d}})$ is trivial due to the fact that $\mathscr{H}_L(\mathscr{W}_{\text{d}})$ has more rows than columns. Therefore, we redefine the notion of most powerful unfalsified model in the finite length case by incorporating the bound $L_{\max} - 1$ on the lag in the definition of $\mathscr{B}_{\text{MPUM}}$.

Definition 12: The most powerful unfalsified model $\mathscr{B} = \mathscr{B}_{\text{MPUM},\mathscr{M}}(\mathscr{W}_{\text{d}})$ for the data $(\mathscr{W}_{\text{d}})$ in the model class $\mathscr{M}$ is defined by the following properties.

1) $\mathscr{B}$ is in the model class and has lag $\boldsymbol{\ell}(\mathscr{B}) < L_{\max}$, i.e.,
$$\mathscr{B} \in \mathscr{M} \cap \mathscr{L}^q_{\bullet, L_{\max}-1}.$$
2) $\mathscr{B}$ is *unfalsified,* i.e.,
$$w_{\text{d}}^i \in \mathscr{B}|_{T_i}, \quad \text{for all } i \in \{ 1, \ldots, N \}. \qquad (\mathscr{W}_{\text{d}} \subset \mathscr{B})$$
3) $\mathscr{B}$ is *most powerful,* i.e., another model $\mathscr{B}'$ for which 1) and 2) hold is no more powerful than $\mathscr{B}$, i.e., $(\mathscr{B} \subseteq \mathscr{B}')$.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:34:11 UTC from IEEE Xplore. Restrictions apply.

Next, we present an algorithm for computing $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_\text{d})$.

Since $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_\text{d})|_L = \text{image}\, \mathscr{H}_L(\mathscr{W}_\text{d})$, for $L = 1, \ldots, L_{\max}$, by Lemma 4 (*cf.*, Note 6) $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_\text{d})$ is uniquely defined by $\mathscr{W}_\text{d}$ as long as its lag is at most $L_{\max}$ (which is enforced by definition). The algorithm constructs sequentially the annihilators of degrees $0, 1, \ldots, L_{\max} - 1$ from the left kernels of

$$\boxed{\mathscr{H}_1(\mathscr{W}_\text{d})}, \quad \boxed{\mathscr{H}_2(\mathscr{W}_\text{d})}, \ldots, \quad \boxed{\mathscr{H}_{L\,\max}(\mathscr{W}_\text{d})}$$

As a result, the algorithm delivers a minimal kernel representation (KER) of $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_\text{d})$. By stopping prematurely when annihilators of degree up to $\ell_{\max} < L_{\max} - 1$ are computed, the resulting set of annihilators define $\mathscr{B}_{\text{MPUM},\mathscr{L}^q_{\bullet,\ell_{\max}}}(\mathscr{W}_\text{d})$. This allows us to easily include a given bound $\ell_{\max}$ on the lag.

The proposed computational method is given in Algorithm 1. Step 4 computes the difference Δr between the predicted rank $Lm + \sum_{i=0}^{p} \ell_i$, based on the currently computed annihilators, and the actual rank of the Hankel matrix $\mathscr{H}_L(\mathscr{W}_\text{d})$. If $\Delta r > 0$, then there are potential new annihilators in the left kernel of $\mathscr{H}_L(\mathscr{W}_\text{d})$. Step 6 computes a basis R_{new} for the subspace spanned by the potential new annihilators and step 7 constructs the potential new annihilators $R^1_{\text{new}}(z), \ldots, R^{\Delta r}_{\text{new}}(z)$ from R_{new}. For a potential new annihilator $R^i_{\text{new}}(z)$ to be an actual annihilator and therefore, be part of the kernel representation of $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_\text{d})$, its degree must be $L - 1$ (or else, it should have been detected on a previous step of the algorithm). Therefore, if $\deg R^i_{\text{new}}(z) = L - 1$, Steps 10 and 11 adapt the model including $R^i_{\text{new}}(z)$ as an annihilator.

The following lemma is a simple but important consequence of Algorithm 1. In particular, its Corollary 14 is used in the proof of the identifiability result Theorem 17 in Section V.

Lemma 13 (The map $\mathscr{B}|_{\ell+1} \mapsto \mathscr{B}$ is well defined): The linear time-invariant system $\mathscr{B} \in \mathscr{L}^{q,n}_{m,\ell}$ is completely specified by its restriction $\mathscr{B}|_{\ell+1}$ to the interval $[1, \ell + 1]$, i.e., there is a unique extension of $\mathscr{B}|_{\ell+1} \subset (\mathbb{R}^q)^{\ell+1}$ to $\mathscr{B} \in \mathscr{L}^{q,n}_{m,\ell}$.

Proof: The proof is constructive. 1) Compute a basis $(\mathscr{W}_\text{d})$ with $T_1 = \cdots = T_N = \ell + 1$ for $\mathscr{B}|_{\ell+1}$, i.e., $(\mathscr{W}_\text{d})$ such that $\text{span}\, \mathscr{W}_\text{d} = \mathscr{B}|_{\ell+1}$ and $N = \dim \mathscr{B}|_{\ell+1} = m(\ell + 1) + n$. 2) For $(\mathscr{W}_\text{d})$, computed on step 1, Algorithm 1 produces a kernel representation of $\mathscr{B}$. This kernel representation can be used then to reconstruct $\mathscr{B}$. □

Corollary 14 ($\mathscr{B}|_{\ell+1} = \mathscr{B}'|_{\ell+1} \iff \mathscr{B} = \mathscr{B}'$): If two systems $\mathscr{B}, \mathscr{B}' \in \mathscr{L}^q$ have the same behaviors over the interval $[1, L]$, where $L \geq \max\{\boldsymbol{\ell}(\mathscr{B}), \boldsymbol{\ell}(\mathscr{B}')\} + 1$, then they coincide.

V. Identifiability

Let the data $(\mathscr{W}_\text{d})$ be generated by a system $\bar{\mathscr{B}} \in \mathscr{M}$, i.e.,

$$w^i_\text{d} \in \bar{\mathscr{B}}|_{T_i}, \quad \text{for all } i \in \{1, \ldots, N\}. \qquad (\mathscr{W}_\text{d} \subset \bar{\mathscr{B}})$$

We refer to $\bar{\mathscr{B}}$ as the *true data-generating system* and pose the following questions: 1) Can we uniquely recover $\bar{\mathscr{B}}$ from $\mathscr{W}_\text{d}$? 2) If so, how can we obtain $\bar{\mathscr{B}}$? Question 1 is called an *identifiability question*. Question 2 is an *exact identification problem*. Identifiability ensures well-posedness of the exact identification problem $\mathscr{W}_\text{d} \mapsto \bar{\mathscr{B}}$.

Algorithm 1: Computation of a Minimal Kernel Representation of the Most Powerful Unfalsified Model $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_\text{d})$.

Input: Data $(\mathscr{W}_\text{d})$.

1: Let $m := q$, $p := 0$, $R^{(0)}(z) := []$, $\ell_0 = 0$. {start with the trivial model $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_\text{d}) = (\mathbb{R}^q)^{\mathbb{N}}$}
2: Let $T := \max\{T_1, \ldots, T_N\}$ and define $(L_{\max})$.
3: **for** $L = 1 : L_{\max}$ **do**
4: Let $\Delta r := Lm + \sum_{i=0}^{p} \ell_i - \text{rank}\, \mathscr{H}_L(\mathscr{W}_\text{d})$. {$\Delta r$ = predicted rank minus actual rank}
5: **if** $\Delta r > 0$ **then**
6: Find potential new annihilators, i.e., R_{new} such that

$$\ker \begin{bmatrix} \mathscr{M}_L(R^{(L-1)}) \\ R_{\text{new}} \end{bmatrix} = \text{left ker}\ \mathscr{H}_L(\mathscr{W}_\text{d}).$$

7: Let $R_{\text{new}} =: \begin{bmatrix} R_{\text{new},0} & R_{\text{new},1} & \cdots & R_{\text{new},L-1} \end{bmatrix}$ and $R_{\text{new}}(z) := R_{\text{new},0} + \cdots + R_{\text{new},L-1} z^{L-1} =: \begin{bmatrix} R^1_{\text{new}}(z) \\ \vdots \\ R^{\Delta r}_{\text{new}}(z) \end{bmatrix}$.
8: **for** $i = 1 : \Delta r$ **do**
9: **if** $R^i_{\text{new},L-1} \neq 0$ {i.e., $\deg R^i_{\text{new}}(z) = L - 1$} **then**
10: Let $m := m - 1$, $p := p + 1$, $\ell_p := L - 1$. {add a new annihilator to the model}
11: Let $R^{(L)}(z) := \begin{bmatrix} R^{(L-1)}(z) \\ R^i_{\text{new}}(z) \end{bmatrix}$. {update the kernel representation with $R^i_{\text{new}}(z)$}
12: **end if**
13: **end for**
14: **end if**
15: **end for**

Output: $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_\text{d}) = \ker R(\sigma)$

Definition 15 (Identifiability): The system $\bar{\mathscr{B}} \in \mathscr{M}$ is identifiable from the data $(\mathscr{W}_\text{d} \subset \bar{\mathscr{B}})$ in the model class $\mathscr{M}$ if $\bar{\mathscr{B}}$ is the only model in $\mathscr{M}$ that fits the data exactly, i.e.,

$$\widehat{\mathscr{B}} \in \mathscr{M} \text{ and } \mathscr{W}_\text{d} \subset \widehat{\mathscr{B}} \quad \Rightarrow \quad \widehat{\mathscr{B}} = \bar{\mathscr{B}}.$$

The identification principle is the Occam's razor—minimization of the complexity over all exact models in a predefined model class $\mathscr{M}$. Using the Occam's razor principle with the prior information that the true model is linear time-invariant, the identifiability question becomes the question of guaranteeing that the most powerful unfalsified model of $\mathscr{W}_\text{d}$ in $\mathscr{L}^q$ coincides with $\bar{\mathscr{B}}$. The following theorems gives conditions on $\mathscr{W}_\text{d}$, under which $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_\text{d}) = \bar{\mathscr{B}}$, so that $\bar{\mathscr{B}}$ is identifiable from $\mathscr{W}_\text{d}$ in $\mathscr{L}^q$.

Theorem 16: Let the data $(\mathscr{W}_\text{d})$ be generated by a system $\bar{\mathscr{B}} \in \partial\mathscr{L}^{q,n}_{m,\ell}$, i.e., $(\mathscr{W}_\text{d} \subset \bar{\mathscr{B}})$. Then, $\bar{\mathscr{B}}$ is identifiable from $\mathscr{W}_\text{d}$, i.e., $\mathscr{B}_{MPUM}(\mathscr{W}_\text{d}) = \bar{\mathscr{B}}$, if with $T := \max\{T_1, \ldots, T_N\}$,

$$\ell < L_{\max} := \left\lfloor \tfrac{T+1}{q+1} \right\rfloor \quad \text{and} \quad \boldsymbol{p}\left(\mathscr{B}_{MPUM}(\mathscr{W}_\text{d})\right) = q - m. \text{ (ID1)}$$

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:34:11 UTC from IEEE Xplore. Restrictions apply.

Proof: The condition $\ell < L_{\max}$ is necessary. Otherwise, there is an annihilator of $\bar{\mathscr{B}}$ with lag larger than or equal to $L_{\max}$ that cannot be detected from the data. Under the assumption $\ell < L_{\max}$, all annihilators of $\bar{\mathscr{B}}$ are detected by Algorithm 1, i.e., they are among the annihilators of $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}})$. However, $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}})$ may include additional annihilators that are not annihilators of $\bar{\mathscr{B}}$. This possibility is ruled out by the condition $\boldsymbol{p}(\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}})) = p = q - m$. Then, the annihilators of $\bar{\mathscr{B}}$ and $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}})$ coincide and therefore, $\bar{\mathscr{B}} = \mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}})$. □

In order to check the identifiability conditions (ID1) of Theorem 16, one has to know the lag ℓ and the number of inputs m of the true system $\bar{\mathscr{B}}$. The conditions (ID1) are sufficient. The following theorem gives a necessary and sufficient condition that requires prior knowledge of the complexity (m, ℓ, n) of $\bar{\mathscr{B}}$.

Theorem 17: Let the data $(\mathscr{W}_{\text{d}})$ be generated by a system $\bar{\mathscr{B}} \in \partial\mathscr{L}^{q,n}_{m,\ell}$, i.e., $(\mathscr{W}_{\text{d}} \subset \bar{\mathscr{B}})$. Then, $\bar{\mathscr{B}}$ is identifiable from $\mathscr{W}_{\text{d}}$, i.e., $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}}) = \bar{\mathscr{B}}$ if and only if

$$\text{rank } \mathscr{H}_{\ell+1}(\mathscr{W}_{\text{d}}) = m(\ell+1) + n. \tag{ID2}$$

Proof: Since the data are exact, $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}}) \subseteq \bar{\mathscr{B}}$. In particular, $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}})|_{\ell+1} \subseteq \bar{\mathscr{B}}|_{\ell+1}$. By (ID2), we have that

$$\dim \mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}})|_{\ell+1} = m(\ell+1) + n.$$

By (4), $\dim \bar{\mathscr{B}}|_{\ell+1} = m(\ell+1) + n$, so that $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}})|_{\ell+1} = \bar{\mathscr{B}}|_{\ell+1}$ and by Corollary 14, $\mathscr{B}_{\text{MPUM}}(\mathscr{W}_{\text{d}}) = \bar{\mathscr{B}}$. □

Theorems 16 and 17 apply to uncontrollable as well as controllable systems and do not require a priori known input/output partitioning of the variables nor controllability. The rank condition (ID2) can be viewed as a generalized persistency of excitation condition for the data $\mathscr{W}_{\text{d}}$.

Note 18 (Comparison with the fundamental lemma [1]): The fundamental lemma gives alternative identifiability conditions:

1) $\bar{\mathscr{B}}$ is controllable;
2) $\bar{\mathscr{B}}$ admits an input/output partitioning $w = \begin{bmatrix} u \\ y \end{bmatrix}$,
3) u_{d} is persistently exciting of order $\ell + n + 1$, i.e.,

$$\text{rank } \mathscr{H}_{\ell+n+1}(u_{\text{d}}) = m(\ell+n+1). \tag{PE}$$

Condition 3 is similar to (ID2): it is a rank condition for a Hankel matrix that requires knowledge of the true model's complexity (m, ℓ, n). The additional Assumptions 1 and 2, however, impose a loss of generality. Although there is always a subset of m variables that are inputs, in general, not *all* subsets of m variables are inputs of $\bar{\mathscr{B}}$, so that the choice of the input/output partition cannot be made without loss of generality. Finally, uncontrollable systems are excluded from consideration. In particular, the fundamental lemma does not include autonomous systems. In contrast, in the case of an autonomous system, Theorem 17 yields the well-known (from realization theory) identifiability condition $\text{rank } \mathscr{H}_{\ell+1}(w_{\text{d}}) = n$.

A noteworthy generalization of the new identifiability results over the fundamental lemma is that they apply to multiple time series. The corresponding generalization of the data structure is from a Hankel matrix to a mosaic-Hankel matrix. Using data from multiple time series has underappreciated potential for system. For example, as illustrated in Section VII, the following corollary involving the trajectory matrix has relevance for approximation in the errors-in-variables setting.

Corollary 19: Let the data $(\mathscr{W}_{\text{d}})$ be generated by a system $\bar{\mathscr{B}} \in \partial\mathscr{L}^{q,n}_{m,\ell}$, i.e., $(\mathscr{W}_{\text{d}} \subset \bar{\mathscr{B}})$ and let $T_i = \ell + 1$ for all $i \in \{1, \ldots, N\}$. Then, $\bar{\mathscr{B}}$ is identifiable from $\mathscr{W}_{\text{d}}$ if and only if

$$\text{rank } \mathscr{T}(\mathscr{W}_{\text{d}}) = m(\ell+1) + n.$$

Note 20: A generalization of the fundamental lemma for multiple time series, involving a mosaic-Hankel matrix, is presented in [3]. The persistency of excitation assumption in [3], however, does not allow for a result similar to Corollary 19 involving the trajectory matrix $\mathscr{T}(\mathscr{W}_{\text{d}})$. An analogous corollary holds for the Page matrix. In [4], an analog to Corollary 19 involving the Page matrix $(\mathscr{P}_L(w))$ is presented under more stringent persistency of excitation assumptions.

The following corollary states that the image of the Hankel matrix $\mathscr{H}_L(w_{\text{d}})$ coincides with the true system's behavior $\bar{\mathscr{B}}|_L$ (i.e., it spans all L-samples long trajectories of $\bar{\mathscr{B}}$).

Corollary 21: Let the data $(\mathscr{W}_{\text{d}})$ be generated by a system $\bar{\mathscr{B}} \in \partial\mathscr{L}^{q,n}_{m,\ell}$, i.e., $(\mathscr{W}_{\text{d}} \subset \bar{\mathscr{B}})$. Then, $\text{image } \mathscr{H}_L(\mathscr{W}) = \bar{\mathscr{B}}_L$, for $L > \ell$ if and only if $\text{rank } \mathscr{H}_L(\mathscr{W}_{\text{d}}) = mL + n$.

Corollary 21 gives a nonparametric data-driven representation. It applies to general multivariable linear time-invariant systems, leads to algorithms, and has been extensively used in data-driven signal processing and control [8]. As shown in the next section modifications of the result allow us to use Page and trajectory matrices in place of the mosaic-Hankel matrix.

VI. Approximation Using Page/trajectory Matrix

The Page matrix $(\mathscr{P}_L(w))$ is introduced in [16] as an alternative to the Hankel matrix $\mathscr{H}_L$ for system realization. Since $\mathscr{P}_L(w_{\text{d}})$ is a submatrix of $\mathscr{H}_L(w_{\text{d}})$

$$\text{image } \mathscr{P}_L(w_{\text{d}}) \subseteq \text{image } \mathscr{H}_L(w_{\text{d}}).$$

Equality does not hold in general. Assuming that it holds, a minimal-order exact model can be obtained either from $\mathscr{P}_L(w_{\text{d}})$ or from $\mathscr{H}_L(w_{\text{d}})$. The models obtained from $\mathscr{P}_L(w_{\text{d}})$ and $\mathscr{H}_L(w_{\text{d}})$ are identical, however, stronger assumptions are needed in order to use $\mathscr{P}_L(w_{\text{d}})$.[1] Therefore, for exact realization there is no advantage of using the Page matrix.

It [16], it is suggested that the Page matrix has advantages when an approximation is needed in case of noisy data. Based on empirical evidence, it is reported that approximate model realization using Kung's method, i.e., unstructured low-rank approximation based on truncation of the singular value decomposition followed by Kalman-Ho realization using the factors of the low-rank approximation gives better results when using the Page matrix in place of the Hankel matrix. The rational for this is that the

[1]A simple necessary condition is that the number of columns of $\mathscr{H}_L(w_{\text{d}})$ and $\mathscr{P}_L(w_{\text{d}})$ are greater than $\dim \bar{\mathscr{B}}|_L = mL + n$, for $L > \ell$. This leads to, $T_{\min} := (m+1)L + n - 1$ samples, for the Hankel matrix, and $T'_{\min} := mL^2 + nL$ samples for the Page matrix: more data is needed for using the Page matrix.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:34:11 UTC from IEEE Xplore. Restrictions apply.

Page matrix $\mathscr{P}_L(w_{\mathrm{d}})$ has no repeated elements and the low-rank approximation $\widehat{D}$ of $\mathscr{P}_L(w_{\mathrm{d}})$, based on truncation of the singular value decomposition, is optimal in the sense of minimizing the Frobenius norm of the approximation error $\mathscr{P}_L(w_{\mathrm{d}}) - \widehat{D}$. Another argument in favor of using the Page matrix in case of noisy data is given in [4], [18], and [25]. In [4], it is shown that the Page matrix is a strictly better distributionally robust predictor of the behavior.

Assuming that w_{d} is an exact trajectory of a linear time-invariant system $\bar{\mathscr{B}} \in \delta\mathscr{L}^{q,n}_{m,\ell}$, both the Page matrix $\mathscr{P}_L(w_{\mathrm{d}})$, for $L > \ell + 1$ and the Hankel matrix $\mathscr{H}_L(w_{\mathrm{d}})$ have *shift-invariant structure* apart from the rank deficiency. For $\mathscr{H}_L(w_{\mathrm{d}})$, the shift-invariant structure manifests itself in the pattern of repeated elements. For $\mathscr{P}_L(w_{\mathrm{d}})$, however, the shift-invariant structure is not evident. The fact that low-rank approximation does not impose shift-invariance of the approximation renders Kung's method using both the Page matrix with $L > \ell + 1$ as well as the Hankel matrix heuristics.

The Page matrix can be viewed alternatively as a trajectory matrix constructed from short segments of a long trajectory. More generally, using multiple experiments instead of a single long experiment the appropriate data matrix is the mosaic Hankel matrix. Using data from multiple experiments has advantages for identification of unstable and uncontrollable systems. For example, multiple experiments allow consistent estimation of an autonomous system [26]. In the special case of using data $\mathscr{W}_{\mathrm{d}}$ consisting of L-samples long trajectories, the mosaic-Hankel matrix $\mathscr{H}_L(\mathscr{W}_{\mathrm{d}})$ becomes the trajectory matrix $\mathscr{T}(\mathscr{W}_{\mathrm{d}})$, i.e., each column of $\mathscr{H}_L(\mathscr{W}_{\mathrm{d}})$ is a different trajectory. Like the Page matrix, the trajectory matrix has no repeated elements. In what follows, we treat the Page matrix and the trajectory matrix together although they use different data (a single long trajectory versus multiple short trajectories).

In the special case $L = \ell + 1$, both $\mathscr{P}_{\ell+1}(w_{\mathrm{d}})$ and $\mathscr{T}(\mathscr{W}_{\mathrm{d}})$ lose the shift-invariance structure. This renders unstructured low-rank approximation of $\mathscr{P}_{\ell+1}(w_{\mathrm{d}})$ and $\mathscr{T}(\mathscr{W}_{\mathrm{d}})$ statistically optimal (ML) for model identification from the corresponding data: multiple short trajectories obtained as segments from w_{d} or independent experiments $\mathscr{W}_{\mathrm{d}}$.

Consider two noisy datasets: 1) $w_{\mathrm{d}} \in (\mathbb{R}^q)^T$ and 2) $(\mathscr{W}_{\mathrm{d}})$ with $T_1 = \cdots = T_N = \ell + 1$ that have the same total number of samples, i.e., $T = N(\ell + 1)$. Although low-rank approximation of $\mathscr{T}(\mathscr{W}_{\mathrm{d}})$ yields a ML estimator using the data $\mathscr{W}_{\mathrm{d}}$, it is not evident that this estimator has superior performance than low-rank approximation of the Hankel matrix $\mathscr{H}_{\ell+1}(w_{\mathrm{d}})$. Indeed, avoiding the shift-invariant structure by using multiple $(\ell + 1)$-long trajectories results in an ℓ-times reduction in the number of columns, which also affects the approximation accuracy. Empirical comparison of low-rank approximation using the Hankel, Page, and trajectory matrices in the errors-in-variance setting is presented next.

VII. Simulation Examples With Noisy Data

In this section, we compare empirically approximate system identification methods based on low-rank approximation of the Hankel matrix, the Page matrix, and the trajectory matrix.

A. Simulation Setup

We use the benchmark example of [27], which is a $n = 4$th order single-input–single-output system $\bar{\mathscr{B}}$ defined by an input/output representation (I/O) with

$$\bar{Q}(z) = 0.50666 + 0.28261z, \quad \text{and}$$

$$\bar{P}(z) = 0.88642 - 1.31608z + 1.58939z^2 - 1.41833z^3 + z^4.$$

Equivalently, the system is defined by a kernel representation (KER) with a parameter

$$\bar{R}(z) = \underbrace{\begin{bmatrix} -\bar{Q}_0 & \bar{P}_0 \end{bmatrix}}_{\bar{R}_0} + \underbrace{\begin{bmatrix} -\bar{Q}_1 & \bar{P}_1 \end{bmatrix}}_{\bar{R}_1} z + \cdots + \underbrace{\begin{bmatrix} -\bar{Q}_4 & \bar{P}_4 \end{bmatrix}}_{\bar{R}_4} z^4.$$

The normalization $\bar{P}_4 = 1$ is used in order to make the parameter vector $\bar{R} := [\bar{R}_0 \quad \bar{R}_1 \quad \bar{R}_2 \quad \bar{R}_3 \quad \bar{R}_4]$ unique.

Two different datasets are generated in the errors-in-variables setup as follows.

1) A single T-samples long trajectory w_{d}.
2) N, $(\ell + 1)$-samples long trajectories $(\mathscr{W}_{\mathrm{d}})$.

In the case of a single trajectory, $w_{\mathrm{d}} = \bar{w} + \tilde{w}$, where $\bar{w} \in \bar{\mathscr{B}}|_T$ is a random trajectory of the system $\bar{\mathscr{B}}$ and $\tilde{w}$ is a zero mean white Gaussian noise with standard deviation s that is varied from 0 (exact data) to 0.1. In the case of multiple trajectories, the noise parameters are identical. The total number of samples in the datasets is the same, i.e., $T = N(\ell + 1)$.

B. Identification Methods

The identification methods compared compute an estimate $\widehat{R}$ of the model parameter $\bar{R}$ from the approximate left kernel of a data matrix: the Hankel matrix $D = \mathscr{H}_{\ell+1}(w_{\mathrm{d}})$, the Page matrix $D = \mathscr{P}_{\ell+1}(w_{\mathrm{d}})$, and the trajectory matrix $D = \mathscr{T}(\mathscr{W})$. The approximation of the left kernel is obtained from the optimal in the Frobenius norm unstructured rank-$m(\ell + 1) + n$ approximation of D, computed by truncation of the singular value decomposition of D. The computed parameter estimate $\widehat{R}$ is normalized. In the SISO case, the normalization is $\widehat{R} := \widehat{R}/\widehat{R}_{\ell+1}$, i.e., we make $\widehat{P}_{\ell+1} = 1$.

The methods based on the singular value decomposition of the Hankel matrix are heuristic. A statistically optimal [maximum likelihood (ML)] estimator in the errors-in-variables setup is obtained by structured low-rank approximation of the Hankel matrix $\mathscr{H}_{\ell+1}(w_{\mathrm{d}})$ [9]. For comparison in the simulation example, we include also the results of the ML estimator, computed by the SLRA package [20], [28], which uses local optimization methods.

C. Validation Criteria

As a validation criterion used is the relative parameter error $e_{\bar{R}} := \|\bar{R} - \widehat{R}\|/\|\bar{R}\|$ averaged over 500 Monte-Carlo repetitions of the estimation with different noise realizations. In addition, we show the error of approximation of the data $\mathscr{W}_{\mathrm{d}}$ by a given model $\mathscr{B}$

$$e_{\mathscr{W}_{\mathrm{d}}} := \sqrt{\textstyle\sum_{i=1}^{N} \min_{\widehat{w}^i \in \mathscr{B}|_{T_i}} \|w^i_{\mathrm{d}} - \widehat{w}^i\|_2^2}.$$

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:34:11 UTC from IEEE Xplore. Restrictions apply.

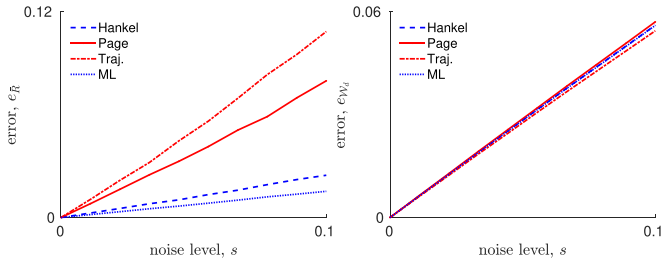


Fig. 2. Low-rank approximation of the trajectory matrix $\mathscr{T}(\mathscr{W}_d)$ (dashed-dotted line) is optimal in the data approximation criterion $e_{\mathscr{W}_d}$, however, it is suboptimal with respect to the parameter estimation error $e_{\bar{R}}$ and yields worse results than low-rank approximation of the Hankel matrix $\mathscr{H}_{\ell+1}(w_d)$ (dashed line). The statistically optimal maximum likelihood (ML) estimate (dotted line) is computed by structured low-rank approximation of $\mathscr{H}_{\ell+1}(w_d)$.

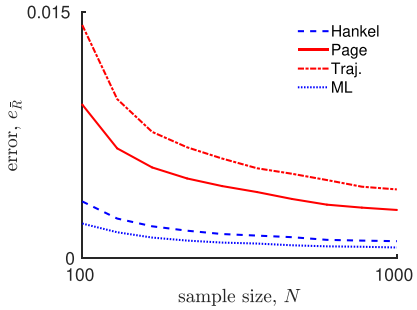


Fig. 3. Estimation errors of the methods converge to zero as the total number of samples goes to infinity (consistency), however, at different rates (efficiency). The efficiency of the low-rank approximation of the Hankel matrix $\mathscr{H}_{\ell+1}(w_d)$ (dashed line) is better than the low-rank approximation of the trajectory matrix $\mathscr{T}(\mathscr{W}_d)$ (dashed-dotted line). Optimal efficiency has the maximum likelihood (ML) estimator (dotted line), computed by structured low-rank approximation of $\mathscr{H}_{\ell+1}(w_d)$.

D. Results and Discussion

Fig. 2 shows the parameter error $e_{\bar{R}}$ and the error of approximation of the data $\mathscr{W}_d$. Low-rank approximation of the trajectory matrix is optimal in terms of the $e_{\mathscr{W}_d}$ criterion. Indeed, $e_{\mathscr{W}_d}$ is the criterion that this method minimizes. With respect to the parameter error $e_{\bar{R}}$, however, low-rank approximation of the Hankel matrix $\mathscr{H}_{\ell+1}(w_d)$ gives better results than low-rank approximation of the Page matrix and the trajectory matrix. This apparent contradiction can be explained by the fact that $\operatorname{col\,dim}\, \mathscr{T}(\mathscr{W}_d) = \operatorname{col\,dim}\, \mathscr{P}_{\ell+1}(w_d) = 100$, while $\operatorname{col\,dim}\, \mathscr{H}_{\ell+1}(w_d) = 495$. Consistency of the estimators corresponding to the methods compared imply that the more columns the data matrix has, the smaller the estimation error $e_{\bar{R}}$ is, as shown in Fig. 3.

The empirical results suggest that although all methods yield consistent estimators, the one based on low-rank approximation of the Hankel matrix is more efficient than the ones based on the Page and trajectory matrices. This is due to the fact that although w_d and $\mathscr{W}_d$ have the same number of samples, w_d is more "informative" for identification of the true system than $\mathscr{W}_d$ and the Hankel matrix exploits this information more effectively than the Page matrix.

VIII. Conclusion

Using formula ($\mathscr{B}|_L$, KER), which is a finite dimensional matrix representation of the restriction of the behavior, we derived an explicit formula for the dimension of the behavior restricted to an interval of any length (Lemma 4). This result is the basis for the other results in the article—definition of model complexity, revision of the notion of most powerful unfalsified model for finite data, and identifiability conditions.

We showed that the classical notion of the most powerful unfalsified model proposed for infinite time series is inadequate in case of finite time series. Irrespective of the data it results in a trivial autonomous model (Lemma 11). We proposed a natural modification of the most powerful unfalsified model for the case of finite time series (Definition 12) that avoids the trivial model without using *priori* knowledge about the model's complexity. This led us to a constructive algorithm (Algorithm 1) for computation of the most powerful unfalsified model. The algorithm computes recursively a minimal kernel representation of the model. The key computational step is detecting rank deficiency of generalized Hankel matrices with increasing depths constructed from the data.

Assuming that the data are generated by a linear time-invariant system of bounded complexity, Theorem 17 gives a necessary and sufficient condition for the most powerful unfalsified model to coincide with the data-generating system. This condition does not require a priori known input/output partitioning and controllability of the true system. It uses only the prior knowledge about the complexity of the true system.

We did not delve into numerical computation issues related to the implementation of Algorithm 1. This topic has connections with work on numerical linear algebra methods for Hankel structured matrices, see for example [29], [30], [31]. In particular incorporating approximation in Algorithm 1 is a interesting topic for further work.

Using a trajectory matrix with $\ell + 1$ block-rows leads to an unstructured data matrix, so that in this case approximation by truncation of the singular value decomposition yields optimal approximation. This observation allows us to avoid the nonconvex optimization of the structured low-rank approximation problem, however, data from multiple short experiments is less informative than data from one long experiment with the same total number of samples (assuming that in both cases the input is a random process with the same distribution) [26]. Empirical results suggest that low-rank approximation of the Hankel matrix gives more accurate model parameter estimates than low-rank approximation of the trajectory matrix. Further research is needed in order to confirm or reject the hypothesis suggested by the empirical results.

Applications of Corollary 21 are interpolation of trajectories, data-driven smoothing, and missing data estimation (see [32]) as well as fault-tolerant control (see [33]).

References

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

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:34:11 UTC from IEEE Xplore. Restrictions apply.

[2] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1999.
[3] H. van Waarde, C. De Persis, M. K. Camlibel, and P. Tesi, "Willems' fundamental lemma for state-space systems and its extension to multiple datasets," *IEEE Control Syst. Lett.*, vol. 4, no. 3, pp. 602–607, Jul. 2020.
[4] J. Coulson, J. Lygeros, and F. Dörfler, "Distributionally robust chance constrained data-enabled predictive control," *IEEE Trans. Automat. Control*, vol. 67, no. 7, pp. 3289–3304, Jul. 2022.
[5] V. Mishra, I. Markovsky, and B. Grossmann, "Data-driven tests for controllability," *Control Syst. Lett.*, vol. 5, pp. 517–522, 2020.
[6] I. Markovsky, J. C. Willems, P. Rapisarda, and B. De Moor, "Algorithms for deterministic balanced subspace identification," *Automatica*, vol. 41, no. 5, pp. 755–766, 2005.
[7] I. Markovsky and P. Rapisarda, "Data-driven simulation and control," *Int. J. Control*, vol. 81, no. 12, pp. 1946–1959, 2008.
[8] 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.
[9] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, no. 4, pp. 891–909, 2008.
[10] I. Markovsky, *Low-Rank Approximation: Algorithms, Implementation, Applications*, 2nd ed. Berlin, Germany: Springer, 2019.
[11] T. Söderström, "Errors-in-variables methods in system identification," *Automatica*, vol. 43, pp. 939–958, 2007.
[12] J. C. Willems, "From time series to linear system–Part II. Exact modelling," *Automatica*, vol. 22, no. 6, pp. 675–694, 1986.
[13] T. Maupong and P. Rapisarda, "Data-driven control: A behavioral approach," *Syst. Control Lett.*, vol. 101, pp. 37–43, 2017.
[14] H. van Waarde, J. Eising, H. Trentelman, and K. Camlibel, "Data informativity: A new perspective on data-driven analysis and control," *IEEE Trans. Automat. Control*, vol. 65, no. 11, pp. 4753–4768, Nov. 2020.
[15] P. Rapisarda and J. C. Willems, "State maps for linear systems," *SIAM J. Control Optim.*, vol. 35, no. 3, pp. 1053–1091, 1997.
[16] A. Damen, P. Van den Hof, and A. Hajdasinski, "Approximate realization based upon an alternative to the Hankel matrix: The page matrix," *Control Syst. Lett.*, vol. 2, pp. 202–208, 1982.
[17] I. Markovsky, J. Goos, K. Usevich, and R. Pintelon, "Realization and identification of autonomous linear periodically time-varying systems," *Automatica*, vol. 50, pp. 1632–1640, 2014.
[18] A. Agarwal, J. Amjad, D. Shah, and D. Shen, "Model agnostic time series analysis via matrix estimation," *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 2, no. 3, pp. 1–39, 2018.
[19] G. Heinig, "Generalized inverses of Hankel and Toeplitz mosaic matrices," *Linear Algebra Appl.*, vol. 216, pp. 43–59, 1995.
[20] I. Markovsky and K. Usevich, "Software for weighted structured low-rank approximation," *J. Comput. Appl. Math.*, vol. 256, pp. 278–292, 2014.
[21] S. Brunton, J. Proctor, and N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proc. Nat. Acad. Sci.*, vol. 113, pp. 3932–3937, 2016.
[22] J. C. Willems, "Paradigms and puzzles in the theory of dynamical systems," *IEEE Trans. Automat. Control.*, vol. 36, no. 3, pp. 259–294, Mar. 1991.
[23] 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. 23, pp. 87–115, 1987.
[24] V. Mishra and I. Markovsky, "The set of linear time-invariant unfalsified models with bounded complexity is affine," *IEEE Trans. Automat. Control*, vol. 66, no. 9, pp. 4432–4435, Sep. 2021.
[25] L. Huang, J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control for grid-connected power converters," in *Proc. IEEE Conf. Decis. Control*, 2019, pp. 8130–8135.
[26] I. Markovsky and R. Pintelon, "Identification of linear time-invariant systems from multiple experiments," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3549–3554, Jul. 2015.
[27] I. Landau, D. Rey, A. Karimi, A. Voda, and A. Franco, "A flexible transmission system as a benchmark for robust digital control," *Eur. J. Control*, vol. 1, no. 2, pp. 77–96, 1995.
[28] I. Markovsky, "A software package for system identification in the behavioral setting," *Control Eng. Pract.*, vol. 21, pp. 1422–1436, 2013.
[29] 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.
[30] G. Heinig and K. Rost, *Algebraic Methods for Toeplitz-Like Matricies and Operators*. Cambridge, MA, USA: Birkhäuser, 1984.
[31] T. Laudadio, N. Mastronardi, and P. Van Dooren, "The generalized Schur algorithm and some applications," *Axioms*, vol. 7, 2018, Art. no. 81.
[32] I. Markovsky and F. Dörfler, "Data-driven dynamic interpolation and approximation," *Automatica*, vol. 135, 2022, Art. no. 110008.
[33] T. Jain and J. Y. D. Sauter, *Active Fault-Tolerant Control Systems: A Behavioral System Theoretic Perspective*. Berlin, Germany: Springer, 2018.



Ivan Markovsky received the Ph.D. degree in electrical engineering from the Katholieke Universiteit Leuven, Leuven, Belgium, in February 2005.

He is currently an ICREA Professor with the International Centre for Numerical Methods in Engineering, Barcelona. 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 are 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.



Florian Dörfler (Senior Member, IEEE) received the Ph.D. degree in mechanical engineering from the University of California at Santa Barbara, Santa Barbara, CA, USA, in 2013, and the diploma degree in engineering cybernetics from the University of Stuttgart, in 2008.

He is currently an Associate Professor with the Automatic Control Laboratory, ETH Zürich, Zürich, Switzerland, and the Associate Head with the Department of Information Technology and Electrical Engineering. From 2013 to 2014, he was an Assistant Professor with the University of California, Los Angeles, Los Angeles, CA, USA. His research interests include control, optimization, and system theory with applications in network systems, in particular electric power grids.

Dr. Dörfler was the recipient of the distinguished young research awards by IFAC (Manfred Thoma Medal 2020) and European Control Award 2020. His students were winners or finalists for Best Student Paper awards at the European Control Conference (2013, 2019), the American Control Conference (2016), the Conference on Decision and Control (2020), the PES General Meeting (2020), the PES PowerTech Conference (2017), and the International Conference on Intelligent Transportation Systems (2021). He was also the recipient of the 2010 ACC Student Best Paper Award, the 2011 O. Hugo Schuck Best Paper Award, the 2012–2014 Automatica Best Paper Award, the 2016 IEEE Circuits and Systems Guillemin-Cauer Best Paper Award, and the 2015 UCSB ME Best PhD award.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on April 04,2025 at 08:34:11 UTC from IEEE Xplore. Restrictions apply.