How to Represent and Identify Affine Time-Invariant Systems?

Ivan Markovsky[a], Jaap Eising[b], and Alberto Padoan[c]

***Abstract*— Affine systems are ubiquitous in modeling and emerge naturally from the linearization of nonlinear dynamics. Despite their relevance in applications, their identification remains largely ad hoc, relying on centering the data before applying linear identification methods. This heuristic approach assumes constant offset and can introduce bias. We develop a dedicated framework for affine system identification, deriving identifiability conditions and identification methods based on difference equation representations. Unlike the classical two-step approach, our method identifies the data-generating system under conditions verifiable from data and system complexity. For noisy data in the errors-in-variables setting, we recast the problem as a structured low-rank approximation, leveraging existing optimization techniques for efficient computation.**

***Index Terms*— System identification, behavioral approach, affine systems, low-rank approximation.**

I. Introduction

In practice, most systems modeled as linear are in fact affine. For instance, linearization of nonlinear systems generically take the form of an affine dynamical systems. In a stochastic setting, dealing with measurement noises and disturbances that are not zero-mean also leads to affine systems. Another generic application area of affine systems is time series analysis, where the data includes an offset. Next, we describe two examples of affine systems coming from the field of dynamic measurements.

- *Heat exchange* Consider the heat exchange process when measuring temperature with a thermometer. Let $w(t)$ be the thermometer's reading at time t and $\bar{w}$ be the environmental temperature. The exchange of heat between the environment and the thermometer is described by the Newton's law of cooling $\frac{d}{dt}w(t) = \alpha\big(\bar{w} - w(t)\big)$, for $t \geq 0$, where $\alpha > 0$ is a constant that depends on the environment and the thermometer.
- *Mass-spring-damper system* As another example, consider measuring a mass M by a scale, which is modeled as a mass-spring-damper system, see Figure 1. The equation describing the position w of the platform is

$$(M+m)\frac{d^2}{dt^2}w(t) = -kw(t) - d\frac{d}{dt}w(t) + M\gamma, \text{ for } t \geq 0,$$

where γ is the gravitational constant.

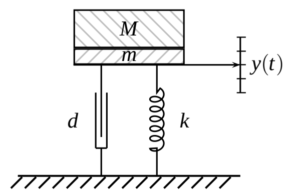

Fig. 1. Weight measurement by a scale, modeled as a mass-spring-damper system, leads to an affine time-invariant system.

In the examples, the systems are defined by affine constant-coefficients differential equations. In the heat exchange example, the constant term in the equation is due to the environmental temperature $\bar{w}$. In the mass-spring-damper system, the constant term is due to the gravity. In both cases, the affine system can be converted to a linear system by redefining the variable w (as $\bar{w} - w$ in the heat exchange example and $\gamma M/k - w$ in the mass-spring-damper system). For analysis of a given system, the transformations can be performed, however, in the metrology application the environmental temperature $\bar{w}$ and the mass M are unknown (to-be-estimated) parameters. This requires modeling the measurement process as an affine system. More generally, moving the equilibrium of a nonlinear system to the origin is not possible when the system is unknown as in system identification and data-driven control.

Since affine systems are ubiquitous in practice, the question of how to identify them from data is important. This question however is often treated in the system identification literature indirectly under "practical aspects" and "data pre-preprocessing", see, *e.g.*, [3, Section 14.1], [4, Section 12.3], and the guide for practitioners [5], which has a dedicated section "Handling Offsets and Trends in Data". The approach recommended in these references is to reduce the affine identification problem to a linear one by *centering the data:*

1) compute the mean of the data and
2) identify a linear model from the centered data.

This largely accepted heuristic, which we refer to as the *two-step procedure*, turns out to have a number of issues. In general, it introduces a bias and can yield arbitrarily suboptimal identification results. In [5], it is noted that preprocessing by mean subtraction should be used for "stationary data."

The heuristic nature of the two-step procedure is also pointed out in [6, Section 10.3.2], where a state-space representation for affine systems is presented, and [7, Section 8.1], where a structured low-rank approximation method

[a] Catalan Institution for Research and Advanced Studies and the Centre Internacional de Mètodes Numèrics en Enginyeria; imarkovsky@cimne.upc.edu
[b] Automatic Control Laboratory, ETH-Zürich; jeising@ethz.ch
[c] Department of Electrical & Computer Engineering, University of British Columbia; apadoan@ece.ubc.ca

is proposed for identification of autonomous systems. The AutoRegressive Integrated Moving Average with eXogenous variables (ARIMAX) model used in econometrics can also deal with constant offsets in the data [8]. It is closely related to the difference equation representation (DE) in this paper, the existence of which we prove starting from a definition of the system as an affine time-invariant set of trajectories.

The goal of this paper is to develop identification methods for *affine time-invariant (ATI)* systems in the general case of "transient data". This is done in the behavioral setting, where we formally consider the class of ATI systems. Then, we define exact and approximate identification problems. The exact identification problem is the problem of recovering the true data-generating system from a trajectory of the system. The approximate identification problem considered is minimization of the distance from the data to the system subject to a bound on the complexity of the system.

Our main contributions are as follows.

- Section II defines the class of ATI systems, its complexity, and three representations—difference equation, input/state/output, and data-driven.
- Section III-A presents an identifiability condition and a method for exact identification. Unlike the classical method based on centering of the data, the method proposed identifies the data-generating system exactly under a condition that is verifiable directly in terms of the data and the true system's complexity condition.
- Section III-B considers the problem of approximate identification of ATI systems, which is reformulated as a structured low-rank approximation problem. The reformulation allows us to use existing methods for solving the problem [7].

II. Preliminaries

We use the behavioral approach to systems theory which views dynamical systems as sets of trajectories [9], [10]. First, we define an ATI behavior. Then, we consider three representations of ATI systems—difference equation, input/state/output, and data-driven.

A. Affine Time-Invariant (ATI) Systems

We consider discrete-time dynamical systems $\mathscr{B} \subset (\mathbb{R}^q)^{\mathbb{N}}$, where $(\mathbb{R}^q)^{\mathbb{N}}$ denotes the set of real q-variate vector signals $\mathbb{N} \to \mathbb{R}^q$. A linear system $\mathscr{B}$ is a linear subspace, *i.e.*, a set $\mathscr{B}$, such that for any two trajectories w, v of $\mathscr{B}$ and any scalars $\alpha, \beta \in \mathbb{R}$, the trajectory $\alpha w + \beta v$ is in $\mathscr{B}$. Similarly, an affine system is defined as an affine set, *i.e.*, a set $\mathscr{B}$, such that for any two trajectories w, v of $\mathscr{B}$ and any scalars $\alpha, \beta \in \mathbb{R}$, with $\alpha + \beta = 1$, the trajectory $\alpha w + \beta v$ is in $\mathscr{B}$.

In this paper, we consider discrete-time systems and denote by σ the unit-shift operator $(\sigma w)(t) := w(t+1)$. Acting on a system $\mathscr{B}$, σ shifts all trajectories $\sigma\mathscr{B} := \{\, \sigma w \mid w \in \mathscr{B} \,\}$. The system $\mathscr{B}$ is time-invariant if $\sigma\mathscr{B} = \mathscr{B}$, when the time axis $\mathbb{Z}$, or $\sigma\mathscr{B} \subseteq \mathscr{B}$, when the time axis is $\mathbb{N}$. The classes of LTI and ATI systems with q variables are denoted by $\mathscr{L}^q$ and $\mathscr{A}^q$, respectively.

An ATI system $\mathscr{B}$ is a "shifted" LTI system $\mathscr{B}_{\text{lin}}$, *i.e.*,

$$\begin{aligned}\mathscr{B} &= \mathscr{B}_{\text{lin}} + w_{\text{c}} \\ &:= \{\, w_{\text{lin}} + w_{\text{c}} \mid w_{\text{lin}} \in \mathscr{B}_{\text{lin}} \,\}, \quad \text{for any } w_{\text{c}} \in \mathscr{B}. \end{aligned} \tag{AFF}$$

The following theorem shows that

$$\mathscr{B}_{\text{lin}} = \mathscr{B} - \mathscr{B} \;:= \{\, w^1 - w^2 \mid w^1, w^2 \in \mathscr{B} \,\}. \tag{$\mathscr{B}_{\text{lin}}$}$$

Theorem 2.1: Let $\mathscr{B} \in \mathscr{A}^q$. Then, there is a unique $\mathscr{B}_{\text{lin}} \in \mathscr{L}^q$, such that (AFF) and ($\mathscr{B}_{\text{lin}}$) hold.

Proof: Since $\mathscr{B}$ is an affine space, there is a unique linear space $\mathscr{B}_{\text{lin}}$, such that $\mathscr{B} = \mathscr{B}_{\text{lin}} + w_{\text{c}}$, for any $w_{\text{c}} \in \mathscr{B}$. By the time-invariance of $\mathscr{B}$, $\sigma w_{\text{c}} \in \mathscr{B}$. Therefore,

$$\sigma w_{\text{c}} - w_{\text{c}} \in \mathscr{B}_{\text{lin}}. \tag{$*$}$$

Consider a trajectory $w \in \mathscr{B}$ and define $w_{\text{lin}} := w - w_{\text{c}} \in \mathscr{B}_{\text{lin}}$. By the time-invariance of $\mathscr{B}$, $\sigma w \in \mathscr{B}$, so that

$$\sigma w - w_{\text{c}} \in \mathscr{B}_{\text{lin}}. \tag{$**$}$$

Subtracting the trajectory in ($*$) from the one in ($**$), we have $\sigma w_{\text{lin}} \in \mathscr{B}_{\text{lin}}$, which proves the time-invariance of $\mathscr{B}_{\text{lin}}$.

In order to prove ($\mathscr{B}_{\text{lin}}$), define $\mathscr{B}'_{\text{lin}} := \sigma\mathscr{B} - \mathscr{B}$ and note that by the shift-invariance $\mathscr{B}'_{\text{lin}} = \mathscr{B} - \mathscr{B}$, which by (AFF) is equal to $\mathscr{B}_{\text{lin}} - \mathscr{B}_{\text{lin}}$, which is equal to $\mathscr{B}_{\text{lin}}$. ∎

Theorem 2.1 shows that for a given ATI system $\mathscr{B}$, the LTI system $\mathscr{B}_{\text{lin}}$ in (AFF) is uniquely defined. The integer invariants: number of inputs $\boldsymbol{m}(\mathscr{B}_{\text{lin}})$, lag $\boldsymbol{\ell}(\mathscr{B}_{\text{lin}})$, and order $\boldsymbol{n}(\mathscr{B}_{\text{lin}})$ are therefore also well-defined. We define the number of inputs $\boldsymbol{m}(\mathscr{B})$, lag $\boldsymbol{\ell}(\mathscr{B})$, and order $\boldsymbol{n}(\mathscr{B})$ of the ATI $\mathscr{B}$ as the corresponding integer invariants of the LTI system $\mathscr{B}_{\text{lin}}$. The *complexity* of $\mathscr{B} \in \mathscr{A}^q$ is defined as

$$\boldsymbol{c}(\mathscr{B}) := \big(\boldsymbol{m}(\mathscr{B}), \boldsymbol{\ell}(\mathscr{B}), \boldsymbol{n}(\mathscr{B})\big).$$

Complexities are ordered lexicographically and the class of ATI systems with bounded complexity is denoted by $\mathscr{A}^q_{(m,\ell,n)}$.

The trajectory w_{c} in (AFF), referred to as an *offset*, is not unique. However, w_{c} must satisfy the constraint $w_{\text{c}} \in \mathscr{B}$ so that it is not free. In practice, of interest is to choose a constant offset: $w_{\text{c}}(t) = \overline{w}_{\text{c}} \in \mathbb{R}^q$ for all t. A constant offset is finitely parametrized and simplifies the derivation of a state-space representation of $\mathscr{B}$ (see Theorem 2.4). In linearization of a nonlinear system around an equilibrium point $\overline{w}_{\text{c}}$, the resulting system is ATI and has a constant offset $\overline{w}_{\text{c}}$. The heat exchange and mass-spring-damper examples from the introduction also lead to ATI systems with a constant offset, where $\overline{w}_{\text{c}}$ is related to physical parameters.

Although a constant offset is important in practice and is universally used in the literature, in general, an ATI system may not have a constant offset. Moreover, when a constant offset exists, in general, it is not unique. Theorem 2.3 gives conditions for existence and uniqueness of a constant offset.

B. Difference Equation Representation

Next, we give a difference equation representation of an ATI system $\mathscr{B}$ based on the kernel representation

$$\mathscr{B}_{\text{lin}} = \ker R_{\text{lin}}(\sigma) := \{\, w \mid R_{\text{lin}}(\sigma) w = 0 \,\}.$$

Theorem 2.2: Consider a bounded complexity ATI system $\mathscr{B} \in \mathscr{A}^q_{(m,\ell,n)}$ and let $\mathscr{B}_{\text{lin}} = \ker R_{\text{lin}}(\sigma)$, where $R_{\text{lin}} \in \mathbb{R}^{p\times q}[z]$. There is $R_{\text{c}} \in \mathbb{R}^p$, such that

$$\mathscr{B} = \{\, w \mid R_{\text{lin}}(\sigma)w = R_{\text{c}} \,\}. \tag{DE}$$

Proof: By Theorem 2.1, for $w \in \mathscr{B}$, $\sigma w - w \in \mathscr{B}_{\text{lin}}$. Then,

$$R_{\text{lin}}(\sigma)(\sigma w - w) = (\sigma - 1)R_{\text{lin}}(\sigma)w = 0. \tag{DE'}$$

This proves that $R_{\text{lin}}(\sigma)w$ is a constant signal, *i.e.*, there is $R_{\text{c}} \in \mathbb{R}^p$, such that $R_{\text{lin}}(\sigma)w = R_{\text{c}}$. In order to show that a fixed R_{c} can be chosen for *all* $w \in \mathscr{B}$, note that there is a fixed trajectory $w_{\text{c}} \in \mathscr{B}$, such that *any* $w \in \mathscr{B}$ can be written as $w = w_{\text{lin}} + w_{\text{c}}$, for some $w_{\text{lin}} \in \mathscr{B}_{\text{lin}}$. For this fixed $w_{\text{c}} \in \mathscr{B}$, by (DE'), we have that there is $R_{\text{c}} \in \mathbb{R}^p$, such that $R_{\text{lin}}(\sigma)w_{\text{c}} = R_{\text{c}}$. For this fixed $R_{\text{c}} \in \mathbb{R}^p$ then, we have that for any $w \in \mathscr{B}$, $R_{\text{lin}}(\sigma)w = R_{\text{lin}}(\sigma)w_{\text{c}} = R_{\text{c}}$. ∎

The difference equation representation (DE) of an ATI system can be viewed as a kernel representation

$$\begin{bmatrix} R_{\text{lin}}(\sigma) & -R_{\text{c}} \end{bmatrix} \begin{bmatrix} w \\ 1 \end{bmatrix} = 0$$

for the extended signal $w_{\text{ext}} := \left[\begin{smallmatrix} w \\ 1 \end{smallmatrix}\right]$. This is a special case of a lifting procedure used in modeling nonlinear systems [11]. The lifting operation $w \mapsto w_{\text{ext}}$ is also related to a standard homogeneous coordinate embedding, which embeds a linear space into a projective space [12]. This construction is fundamental in algebraic geometry, where an affine space is associated to a projective space, ensuring that affine transformations are expressed as linear maps.

Theorem 2.3: An ATI system with a representation (DE) has a constant offset $\overline{w}_{\text{c}}$ if and only if $R_{\text{lin}}(1)\overline{w}_{\text{c}} = R_{\text{c}}$ has a solution. The $\overline{w}_{\text{c}}$ is unique if and only if $R_{\text{lin}}(1)$ is full column rank.

Proof: For a constant $w_{\text{c}} \in \mathscr{B}$, *i.e.*, $w_{\text{c}}(t) = \overline{w}_{\text{c}}$ for all t,

$$\big(R_{\text{lin}}(\sigma)w_{\text{c}}\big)(t) = R_{\text{lin}}(1)\overline{w}_{\text{c}}, \quad \text{for all } t.$$

Then, (DE) holds iff $R_{\text{lin}}(1)\overline{w}_{\text{c}} = R_{\text{c}}$ has a solution. ∎

As in the LTI case, partitioning the variables of an ATI system into inputs and outputs leads to the representation

$$\mathscr{B} = \{\, \Pi \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \mid P_{\text{lin}}(\sigma)y = Q_{\text{lin}}(\sigma)u + R_{\text{c}} \,\}, \tag{I/O}$$

where $\Pi \in \mathbb{R}^{q\times q}$ is a permutation matrix, $P_{\text{lin}}(z) \in \mathbb{R}^{p\times p}[z]$, and $Q_{\text{lin}}(z) \in \mathbb{R}^{p\times m}[z]$, with $p := q - m$ and $m := \boldsymbol{m}(\mathscr{B})$. The input/output representation (I/O) is convenient for simulation because it defines a recursive formula for the computation of the output $y(t)$ for a given input $u(t)$ and initial conditions, specified by past samples $w(t-1), \dots, w(t - \deg R_{\text{lin}})$.

In order to simplify the notation, in what follows, we assume that Π is the identity matrix.

C. Input/State/Output Representation

Theorem 2.4: Consider a bounded complexity ATI system $\mathscr{B} \in \mathscr{A}^q_{(m,\ell,n)}$, for which a constant offset w_{c} exists, and let

$$\mathscr{B}_{\text{lin}} = \{\, \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \mid \sigma x = Ax + Bu,\ y = Cx + Du \,\}, \tag{SS-LIN}$$

be a minimal input/state/output representation of $\mathscr{B}_{\text{lin}}$. Then, there are vectors $e \in \mathbb{R}^n$ and $f \in \mathbb{R}^p$, such that

$$\mathscr{B} = \{\, \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \mid \text{there is } x, \text{ such that } \sigma x = Ax + Bu + e, \\ y = Cx + Du + f \,\}. \tag{SS}$$

Proof: By Theorem 2.1, for any $w \in \mathscr{B}$, $w = w_{\text{lin}} + w_{\text{c}}$, where $w_{\text{lin}} \in \mathscr{B}_{\text{lin}}$ and $w_{\text{c}} \in \mathscr{B}$. By assumption, the offset w_{c} can be chosen constant $w_{\text{c}}(t) = \overline{w}_{\text{c}}$ for all t. Substituting $u_{\text{lin}} = u - u_{\text{c}}$ and $y_{\text{lin}} = y - y_{\text{c}}$ in the state and output equations of (SS-LIN), we obtain the state and output equations of (SS) with $e = B\overline{u}_{\text{c}}$ and $f = Du_{\text{c}} - \overline{y}_{\text{c}}$, where $\overline{w}_{\text{c}} =: \left[\begin{smallmatrix} \overline{u}_{\text{c}} \\ \overline{y}_{\text{c}} \end{smallmatrix}\right]$. ∎

D. Data-Driven Representation

Direct data-driven methods for LTI analysis and control are based on a representation of the finite-horizon behavior

$$\mathscr{B}_{\text{lin}}|_L := \{\, w|_L \mid w \in \mathscr{B}_{\text{lin}} \,\}, \quad \text{where } w|_L := \big(w(1), \dots, w(L)\big)$$

by the image of the Hankel matrix

$$\mathscr{H}_L(w_{\text{d}}) := \begin{bmatrix} w_{\text{d}}(1) & w_{\text{d}}(2) & \cdots & w_{\text{d}}(T_{\text{d}} - L + 1) \\ w_{\text{d}}(2) & w_{\text{d}}(3) & \cdots & w_{\text{d}}(T_{\text{d}} - L + 2) \\ \vdots & \vdots & & \vdots \\ w_{\text{d}}(L) & \cdots & \cdots & w_{\text{d}}(T_{\text{d}}) \end{bmatrix},$$

constructed from a trajectory $w_{\text{d}} \in \mathscr{B}_{\text{lin}}|_{T_{\text{d}}}$ of the system [10]. Conditions for

$$\mathscr{B}_{\text{lin}}|_L = \text{image}\, \mathscr{H}_L(w_{\text{d}}) \tag{DDR}$$

to hold true are given in what became known as the fundamental lemma [13, Theorem 1]. The conditions of the fundamental lemma require an input/output partitioning of the variables, persistency of excitation of the input component u_{d} of w_{d} and controllability of the data generating system $\mathscr{B}_{\text{lin}}$. These conditions are suitable for input design as they restrict the input and the data generating system in order they guarantee (DDR) for any initial condition under which the data collection experiment is performed. An alternative necessary and sufficient condition

$$\text{rank}\, \mathscr{H}_L(w_{\text{d}}) = \boldsymbol{m}(\mathscr{B}_{\text{lin}})L + \boldsymbol{n}(\mathscr{B}_{\text{lin}}) \tag{GPE}$$

for the data-driven representation (DDR), called *generalized persistency of excitation*, is given in [14, Corollary 21]. The generalized persistency of excitation requires prior knowledge of the complexity of $\mathscr{B}_{\text{lin}}$ but, unlike the fundamental lemma, no input/output partitioning and no controllability. The generalized persistency of excitation condition is suitable for checking when given w_{d} can represent $\mathscr{B}_{\text{lin}}|_L$.

A fundamental lemma like result for ATI systems is presented in [15], [16]. A result similar to [14, Corollary 21] for ATI systems is stated without a proof in [17]. Next, we prove it. Define the $T \times 1$ vector $\mathbf{1}$ of all ones and the *affine span* of a matrix $H \in \mathbb{R}^{L\times T}$ as $\text{aff}\, H := \{\, Hg \mid \mathbf{1}^\top g = 1 \,\}$.

Lemma 2.5: Let $\mathscr{B} \in \mathscr{A}^q$, $w_{\text{d}} \in \mathscr{B}_{T_{\text{d}}}$, and $L > \boldsymbol{\ell}(\mathscr{B})$. Assume that $\mathscr{B}$ admits a constant offset. Then,

$$\mathscr{B}|_L = \text{aff}\, \mathscr{H}_L(w_{\text{d}}) \tag{ATI-DDR}$$

if and only if

$$\operatorname{rank}\begin{bmatrix}\mathscr{H}_L(w_{\rm d})\\ \mathbf{1}^\top\end{bmatrix} = \boldsymbol{m}(\mathscr{B})L+\boldsymbol{n}(\mathscr{B})+1. \tag{ATI-GPE}$$

Proof: By Theorem 2.4, $\mathscr{B}$ admits a representation (SS). Without loss of generality assume that (SS) is minimal, *i.e.*, its state dimension is $n=\boldsymbol{n}(\mathscr{B})$. Define the extended observability matrix $\mathscr{O}_L$, convolution matrix $\mathscr{T}_L(H)$,

$$\mathscr{O}_L := \begin{bmatrix} C\\ CA\\ \vdots\\ CA^{L-1}\end{bmatrix},\quad \mathscr{T}_L(H) := \begin{bmatrix} H(0) & 0 & \cdots & 0\\ H(1) & H(0) & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ H(L-1) & \cdots & H(1) & H(0)\end{bmatrix},$$

and the Markov parameters

$$H(0) := D,\ H'(0) = I,\ H(t) := CA^{t-1}B,\ H'(t) = CA^{t-1},$$

for $t>0$. For any $\left[\begin{smallmatrix}u\\ y\end{smallmatrix}\right]\in\mathscr{B}|_L$, there is $x_0\in\mathbb{R}^n$, such that

$$\begin{bmatrix}u\\ y\end{bmatrix} = \underbrace{\begin{bmatrix} 0 & I_{mL} & 0\\ \mathscr{O}_L & \mathscr{T}_L(H) & \mathbf{1}\otimes f+\mathscr{T}_L(H')(\mathbf{1}\otimes e)\end{bmatrix}}_{\mathscr{M}_L\in\mathbb{R}^{qL\times(mL+n+1)}}\begin{bmatrix}x_0\\ u\\ 1\end{bmatrix}. \tag{$\mathscr{M}$}$$

(if) Since $\mathscr{B}$ is ATI and all columns of $\mathscr{H}_L(w_{\rm d})$ belong to $\mathscr{B}_L$, aff $\mathscr{H}_L(w_{\rm d})\subseteq\mathscr{B}_L$. Next, we show that every $w\in\mathscr{B}_L$ is in aff $\mathscr{H}_L(w_{\rm d})$. Let $x_{\rm d}$ be the state sequence associated with $w_{\rm d}$. Using ($\mathscr{M}$) and defining $e := \begin{bmatrix}0 & \cdots & 0 & 1\end{bmatrix}$, $w|_L =: \mathbf{\Pi}\left[\begin{smallmatrix}u|_L\\ y|_L\end{smallmatrix}\right]$

$$\begin{bmatrix}\mathscr{H}_L(w_{\rm d})\\ \mathbf{1}^\top\end{bmatrix} = \begin{bmatrix}\mathbf{\Pi}\mathscr{M}_L\\ e\end{bmatrix}\begin{bmatrix}\mathscr{H}_1(x_{\rm d})\\ \mathscr{H}_L(u_{\rm d})\\ \mathbf{1}^\top\end{bmatrix}.$$

Since by the minimality of the state-space representation, $\left[\begin{smallmatrix}\mathbf{\Pi}\mathscr{M}_L\\ e\end{smallmatrix}\right]$ is full column rank,

$$\operatorname{rank}\begin{bmatrix}\mathscr{H}_L(u_{\rm d})\\ \mathbf{1}^\top\end{bmatrix} = \operatorname{rank}\begin{bmatrix}\mathscr{H}_1(x_{\rm d})\\ \mathscr{H}_L(u_{\rm d})\\ \mathbf{1}^\top\end{bmatrix} = mL+n+1.$$

Therefore, for any $(x_0,u,1)\in\mathbb{R}^{mL+n+1}$, there is $g\in\mathbb{R}^{T_{\rm d}-L+1}$, such that $\mathbf{1}^\top g=1$ and $w=\mathscr{H}_L(w_{\rm d})g$, *i.e.*, $w\in$ aff $\mathscr{H}_L(w_{\rm d})$.

(only if) Suppose by contradiction that $\operatorname{rank}\left[\begin{smallmatrix}\mathscr{H}_L(w_{\rm d})\\ \mathbf{1}^\top\end{smallmatrix}\right] < mL+n+1$. Then, there are $x_0\in\mathbb{R}^n$ and $u\in\mathbb{R}^{mL}$, such that

$$\begin{bmatrix}x_0\\ u\\ 1\end{bmatrix} = \begin{bmatrix}\mathscr{H}_1(x_{\rm d})\\ \mathscr{H}_L(u_{\rm d})\\ \mathbf{1}^\top\end{bmatrix} g$$

doesn't hold for any g. Define $\widehat{w} := \mathscr{M}\left[\begin{smallmatrix}x_0\\ u\\ 1\end{smallmatrix}\right]$. By ($\mathscr{M}$), $\widehat{w}\in\mathscr{B}|_L$, however, $\widehat{w}\notin$ aff $\mathscr{H}_L(w_{\rm d})$. This contradicts the assumption that $\mathscr{B}|_L=$ aff $\mathscr{H}(w_{\rm d})$ and thus proves the claim. ∎

In analogy with the LTI case, we call the rank condition (ATI-GPE) *generalized persistence of excitation*.

III. Identification of ATI systems

The *ATI exact identification problem* is defined as follows. Given a trajectory $w_{\rm d}\in\mathscr{B}|_{T_{\rm d}}$ of an ATI system $\mathscr{B}\in\mathscr{A}^q$, find $\mathscr{B}$. When $w_{\rm d}$ is not a trajectory of $\mathscr{B}$, the lack of fit, called *misfit*, between $w_{\rm d}$ and $\mathscr{B}$ is defined as

$$\operatorname{dist}(w_{\rm d},\mathscr{B}) := \min_{\widehat{w}\in\mathscr{B}|_{T_{\rm d}}} \|w_{\rm d}-\widehat{w}\|,$$

with $\|\cdot\|$ denoting the 2-norm. The *ATI approximate identification problem* considered is defined as follows. Given a trajectory $w_{\rm d}\in(\mathbb{R}^q)^{T_{\rm d}}$ and a complexity (m,ℓ,n),

$$\text{minimize}\quad\text{over}\quad \widehat{\mathscr{B}}\in\mathscr{A}^q_{(m,\ell,n)}\quad \operatorname{dist}(w_{\rm d},\widehat{\mathscr{B}}). \tag{P}$$

Problem (P) is the maximum-likelihood identification problem in the errors-in-variables setting [18]

$$w_{\rm d} = \overline{w}_{\rm d}+\widetilde{w}_{\rm d},\quad \text{where } \overline{w}_{\rm d}\in\mathscr{B}|_{T_{\rm d}},\ \mathscr{B}\in\mathscr{A}^q_{(m,\ell,n)}. \tag{EIV}$$

with zero mean, white, Gaussian noise $\widetilde{w}_{\rm d}$.

A. Exact Identification

The first question in solving the ATI exact identification problem is "When does a solution exist?" This is the identifiability question. *Identifiability* is a property of the data $w_{\rm d}$, under the data-generating system $\mathscr{B}$ can be recovered from $w_{\rm d}$. We give an identifiability condition that depends on the data and the true system's complexity. The proof is constructive and leads to a computational method for solving the ATI exact identification problem.

Consider an ATI system $\mathscr{B}\in\mathscr{A}^q_{(m,\ell,n)}$ with a minimal input/state/output representation (SS). Define the LTI system $\mathscr{B}_{\rm aug}\in\mathscr{L}^q_{(m,\ell,n+1)}$ with the state-space representation

$$\begin{aligned}\sigma x_{\rm aug} &= \begin{bmatrix}A & e\\ 0_{1\times n} & 1\end{bmatrix}x_{\rm aug}+\begin{bmatrix}B\\ 0\end{bmatrix}u\\ y &= \begin{bmatrix}C & f\end{bmatrix}x_{\rm aug}+Du.\end{aligned} \tag{SS-AUG}$$

The system $\mathscr{B}_{\rm aug}$, referred to as the *augmented system*, is uniquely defined by $\mathscr{B}$ (*i.e.*, it does not depend on the choice of the representation (SS)). Its importance is due to

$$w\in\mathscr{B}\quad\iff\quad w\in\mathscr{B}_{\rm aug}\ \text{and}\ x_{\rm aug}(0)=\begin{bmatrix}x(0)\\ 1\end{bmatrix}. \tag{$x_{\rm aug}$}$$

It allows us to use LTI methods for simulation, analysis, control of ATI systems. Indeed, the ATI system $\mathscr{B}\in\mathscr{A}^q_{(m,\ell,n)}$ is equivalent to the LTI system $\mathscr{B}_{\rm aug}$ with the constraint ($x_{\rm aug}$) on the initial state. Note also that the order of $\mathscr{B}_{\rm aug}$ is $n+1$ and it has an eigenvalue at 1.

The last state equation in (SS-AUG) together with the initial condition ($x_{\rm aug}$) is a model

$$\sigma x_{{\rm aug},n+1} = x_{{\rm aug},n+1},\ x_{{\rm aug},n+1}(0)=1$$

of a constant signal $x_{{\rm aug},n+1}=1$. Thus, $\mathscr{B}_{\rm aug}$ includes an "internal model" of a constant.

Theorem 3.1: Let $\mathscr{B}\in\mathscr{A}^q$ and $w_{\rm d}\in\mathscr{B}|_{T_{\rm d}}$. Then,

$$\operatorname{rank}\mathscr{H}_L(w_{\rm d})\le\boldsymbol{m}(\mathscr{B})L+\boldsymbol{n}(\mathscr{B})+1,\quad\text{for all } L\ge\boldsymbol{\ell}(\mathscr{B}).$$

The proof follows from $w_{\rm d}\in\mathscr{B}_{\rm aug}|_{T_{\rm d}}$ and [14, Corollary 5].

Theorem 3.2: Let $\mathscr{B}\in\mathscr{A}^q$, $w_{\rm d}\in\mathscr{B}|_{T_{\rm d}}$, and $\ell:=\boldsymbol{\ell}(\mathscr{B})$. Assume that $\mathscr{B}$ admits a constant offset. Then, $\mathscr{B}$ is identifiable from $w_{\rm d}$ and the prior knowledge of $\boldsymbol{\ell}(\mathscr{B})$ if and only if

$$\operatorname{rank}\begin{bmatrix}\mathscr{H}_{\ell+1}(w_{\rm d})\\ \mathbf{1}^\top_{T_{\rm d}-\ell}\end{bmatrix} = \boldsymbol{m}(\mathscr{B})(\ell+1)+\boldsymbol{n}(\mathscr{B})+1. \tag{GPE-IDENT}$$

Proof: By Lemma 2.5, under the (GPE-IDENT) condition, the finite-horizon behavior $\mathscr{B}|_L$ with $L=\ell+1$ is identified from the data by (ATI-DDR). The extension of $\mathscr{B}|_L$ to $\mathscr{B}$ is also well-defined since $L > \boldsymbol{\ell}(\mathscr{B})$. ∎

Consider for example the difference equation representation (DE) of $\mathscr{B}$. Its parameters R_{lin} and R_{c} satisfy

$$\begin{bmatrix} R_{\text{lin}} & R_{\text{c}} \end{bmatrix} \begin{bmatrix} \mathscr{H}_{\ell+1}(w_{\text{d}}) \\ \mathbf{1}^\top_{T_{\text{d}}-\ell} \end{bmatrix} =: RH = 0. \tag{DE''}$$

so that they can be computed from the data. This leads us to the ATI identification method in Algorithm 1.

The condition (GPE-IDENT) will be referred to as the *generalized persistency of excitation condition for identifiability*. Theorem 3.2 states that it is sufficient for identifiability.

Algorithm 1 ATI exact system identification using the difference equation representation (DE).

Input: Data $w_{\text{d}} \in \mathscr{B}|_T$ and lag $\ell = \boldsymbol{\ell}(\mathscr{B})$

1: Construct the matrix H in (DE'').

2: Compute a basis for the left null space of H and let $\widehat{R}$ be the matrix of the basis vectors.

3: Partition $\widehat{R}$ as $\begin{bmatrix} \widehat{R}_{\text{lin}} & \widehat{R}_{\text{c}} \end{bmatrix}$ and let $\widehat{R}_{\text{lin}}(z)$ be the polynomial defined by $\widehat{R}_{\text{lin}}$.

Output: Identified system $\widehat{\mathscr{B}}$ defined by (DE).

The results are trivially generalizable for data

$$\mathscr{W}_{\text{d}} = \{w^1_{\text{d}}, \dots, w^N_{\text{d}}\}, \quad w^i_{\text{d}} \in (\mathbb{R}^q)^{T_{\text{d},i}}.$$

consisting of multiple trajectories of the system by replacing the Hankel matrix $\mathscr{H}_L(w_{\text{d}})$ with the mosaic-Hankel matrix

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

B. Approximate Identification

Algorithm 1 computes model parameters R_{lin} and R_{c} from the left null space of the extended Hankel matrix H, see (DE''). Under the identifiability condition (GPE-IDENT), Algorithm 1 recovers the data generating system. In case of noisy data, generically H is full row rank which implies that there is no model in the model class of bounded complexity $\mathscr{A}_{(m,\ell,n)}$ that fits the data exactly. An adaptation that allows us to recover an approximate model is to replace the null space computation with approximate null space computation by the singular value decomposition (SVD). Alternatively, the generalized SVD [19], taking into account that the last row of $\widehat{H}$ is $\mathbf{1}^\top_{T_{\text{d}}-\ell}$, can be used.

The approximate left null space computation by the (generalized) SVD can be viewed alternatively as a (generalized) low-rank approximation of the matrix H. The approximate left null space obtained from the SVD corresponds to the null space of the closest in the Frobenius norm rank-r matrix $\widehat{H}$ to H. By the Eckart-Schmidt-Mirsky theorem [20], $\widehat{H}$ is obtained by truncation of the SVD of H. In accordance with (GPE), in the modification of Algorithm 1 for approximate identification, we enforce rank $r := m(\ell+1)+n+1$.

The heuristic nature of the modifications of Algorithm 1 by the (generalized) low-rank approximation is due to the fact that $\widehat{H}$ does not preserve the Hankel structure of H. The structure of $\widehat{H}$ is needed in order for (DE'') to correspond to a representation (DE) of a bounded complexity ATI system. Preserving the structure of H, while reducing its rank, leads to the structured low-rank approximation problem

$$\begin{aligned} &\text{minimize} \quad \text{over } \widehat{w} \in (\mathbb{R}^q)^{T_{\text{d}}} \quad \|w_{\text{d}} - \widehat{w}\| \\ &\text{subject to} \quad \operatorname{rank} \begin{bmatrix} \mathscr{H}_{\ell+1}(\widehat{w}) \\ \mathbf{1}^\top_{T_{\text{d}}-\ell} \end{bmatrix} = m(\ell+1)+n+1, \end{aligned} \tag{SLRA}$$

which doesn't admit an analytic solution. In Section IV-B, we compare the low-rank approximation, generalized low-rank approximation, and a local optimization method for (SLRA).

IV. Simulation Examples

Section IV-A shows the minimum-norm and a constant offset of an ATI system and illustrates the results of Section III-A—empirically confirms Theorem 3.2, shows that Algorithm 1 recovers the data-generating system exactly while the classical two-step procedure doesn't, and confirms (SS-AUG) by checking the eigenvalues of $\mathscr{B}_{\text{lin}}$ and $\mathscr{B}_{\text{aug}}$. Section IV-B illustrates the methods for approximate system identification, comparing their performance empirically. All simulation are done in Matlab and the files are available from
`https://imarkovs.github.io/ati-sysid.pdf`

A. Exact identification

Consider the 4th order single-input single-output ($q=2$, $m=1$, $n=4$) ATI system $\mathscr{B}$ defined by (DE) with parameters

$$\begin{aligned} R_{\text{lin}}(z) = \begin{bmatrix} 0.5 & -0.9 \end{bmatrix} z^0 + \begin{bmatrix} 0.3 & 1.3 \end{bmatrix} z^1 + \begin{bmatrix} 0 & -1.6 \end{bmatrix} z^2 \\ + \begin{bmatrix} 0 & 1.4 \end{bmatrix} z^3 + \begin{bmatrix} 0 & -1 \end{bmatrix} z^4 \quad \text{and} \quad R_{\text{c}} = 1. \end{aligned}$$

and three trajectories of $\mathscr{B}$ with length $T_{\text{d}} = 100$:

- w_{c}—constant trajectory,
- w^*_{c}—the min-norm trajectory $w^*_{\text{c}} := \arg\min_{w_{\text{c}} \in \mathscr{B}|_T} \|w_{\text{c}}\|$,
- w_{d}—randomly chosen trajectory.

The constant and minimum-norm trajectories, shown in Figure 2, don't satisfy (GPE-IDENT). The random trajectory w_{d}, selected as $w_{\text{lin}} + w^*_{\text{c}}$ where w_{lin} is a random trajectory of $\mathscr{B}_{\text{lin}}$, satisfies (GPE-IDENT). Using Algorithm 1 and w_{d}, we identify the data-generating system $\mathscr{B}$. In contrast, the two-step approach—center the data and identify $\mathscr{B}_{\text{lin}}$ from the centered data—doesn't recover the true system.

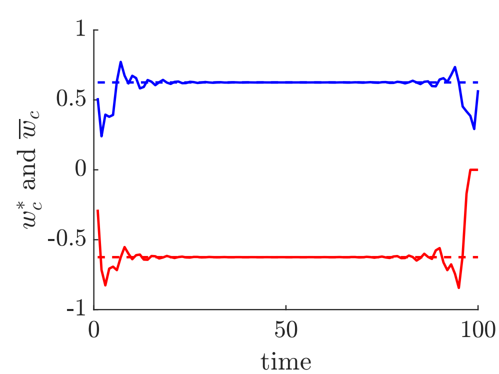

Fig. 2. Minimum-norm (solid line) and constant (dashed line) offsets.

There is an exact model $\mathscr{B}_{\text{aug}}$ for w_{d} in $\mathscr{L}^2_{(1,5,5)}$. We verify that its eigenvalues are the eigenvalues of $\mathscr{B}_{\text{lin}}$ and an extra eigenvalue 1, which empirically confirms (SS-AUG).

B. Approximate identification

This section compares the performance of the methods presented in Section III-B:

- `lra` — Algorithm 1 with low-rank approximation,
- `glra` — Algorithm 1 with generalized low-rank [19],
- `slra` — the method of [21] for solving (SLRA),

and classical two-step approach:

- `2-step` — centering followed by LTI identification (using the N4SID method [22]).

As a test example, we use the mass-spring-damper system from the introduction. Noisy data is generated in accordance with the errors-invariables model (EIV). The validation criterion is the relative normalized parameter $R := \begin{bmatrix} R_c & R_{\text{lin}} \end{bmatrix} / R_c$ estimation error: $e_R(\widehat{R}) := 100\,\|\overline{R} - \widehat{R}\| / \|\overline{R}\|\,\%$, where, $\overline{R}$ is the normalized true system's parameter and $\widehat{R}$ is the normalized estimated system's parameter. The results, averaged over 100 Monte-Carlo simulations are shown in Table I.

TABLE I

std	LRA	GLRA	SLRA	2-step
0	10^{-12}	10^{-12}	10^{-14}	71.036
2.510^{-4}	0.4284	0.0750	0.0024	71.036
5.010^{-4}	1.7611	0.1389	0.0180	71.036
7.510^{-4}	3.9200	0.1948	0.0903	71.036
1.010^{-3}	6.7066	0.2925	0.2826	71.036

Next, we use the identified models for simulation of a trajectory of the system. In addition to the indirect model-based approach using the identified models, we use the data-driven representation (ATI-DDR), modified with low-rank approximation, for direct data-driven simulation of the ATI system's trajectory. The validation criterion is the relative simulation error $e_y(\widehat{y}) := 100\,\|\overline{y} - \widehat{y}\| / \|\overline{y}\|\,\%$, where $\overline{y}$ is the true output and $\widehat{y}$ is the estimated one. The results, averaged over 100 Monte-Carlo simulations, are shown in Table II.

TABLE II

std	DDR	LRA	GLRA	SLRA	2-step
0	10^{-12}	10^{-12}	10^{-11}	10^{-12}	19.3666
2.510^{-4}	0.0590	1.3789	0.2584	0.0075	29.5716
5.010^{-4}	0.1114	5.7853	0.4786	0.0596	29.5813
7.510^{-4}	0.1895	13.2623	0.6752	0.3004	29.6761
1.010^{-3}	0.2732	23.6218	1.0060	0.9322	29.5241

The two-step method fails because it assumes the offset can be consistently estimated via data centering. However, this introduces bias in case of transient data. The simulation uses a transient trajectory of an autonomous system, violating the stationarity assumption underlying the two-step approach.

V. Conclusions

Affine systems naturally arise in system modeling, analysis, and control, particularly through the linearization of nonlinear dynamics, motivating a dedicated study of their structure and identifiability. In this work, we formally framed the identifiability problem of ATI systems within behavioral systems theory and derived identifiability conditions, highlighting the need to adapt the notion of persistence of excitation to this model class. The proposed algorithms, designed for both exact and approximate identification, outperformed the classical two-step method based on data centering. Future work will explore input design and control of affine systems, as well as extension of the results to nonlinear systems.

Acknowledgments

The research leading to these results has received funding from: the Catalan Institution for Research and Advanced Studies (ICREA), the Fond for Scientific Research Vlaanderen (FWO) G033822N, and the Spanish Ministry of Science MCIU/AEI/10.13039/501100011033, PID2023-148952OB-I00.

References

[1] M. Lazar, "Neural data-enabled predictive control," arXiv:2406.08003, 2024.

[2] J. Berberich and F. Allgöwer, "An overview of systems-theoretic guarantees in data-driven model predictive control," *Annual Review of Control, Robotics, and Aut. Sys.*, vol. 8, pp. 77–100, 2025.

[3] L. Ljung, *System Identification: Theory for the User*. Prentice-Hall, 1999.

[4] T. Söderström and P. Stoica, *System Identification*. Prentice Hall, 1989.

[5] L. Ljung, *System Identification Toolbox: User's guide*, The MathWorks, 2023.

[6] M. Verhaegen and V. Verdult, *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.

[7] I. Markovsky, *Low-Rank Approximation: Algorithms, Implementation, Applications*. Springer, 2019.

[8] G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

[9] J. C. Willems, "The behavioral approach to open and interconnected systems: Modeling by tearing, zooming, and linking," *IEEE Control Systems Magazine*, vol. 27, pp. 46–99, 2007.

[10] I. Markovsky, L. Huang, and F. Dörfler, "Data-driven control based on behavioral approach: From theory to applications in power systems," *IEEE Control Systems Magazine*, vol. 43, pp. 28–68, 2023.

[11] I. Markovsky, "Data-driven simulation of generalized bilinear systems via linear time-invariant embedding," *IEEE Trans. Automat. Contr.*, vol. 68, pp. 1101–1106, 2023.

[12] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms*. Springer, 2015.

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

[14] I. Markovsky and F. Dörfler, "Identifiability in the behavioral setting," *IEEE Trans. Automat. Contr.*, vol. 68, pp. 1667–1677, 2023.

[15] J. Berberich, J. Köhler, M. A. Müller, and F. Allgöwer, "Linear tracking MPC for nonlinear systems—part II: The data-driven case," *IEEE Trans. Automat. Contr.*, vol. 67, no. 9, pp. 4406–4421, 2022.

[16] A. Martinelli, M. Gargiani, M. Draskovic, and J. Lygeros, "Data-driven optimal control of affine systems: A linear programming perspective," *IEEE Control Systems Letters*, vol. 6, pp. 3092–3097, 2022.

[17] A. Padoan, F. Dörfler, and J. Lygeros, "Data-driven representations of conical, convex, and affine behaviors," in *Proc. of the IEEE Conference on Decision and Control*, 2023, pp. 596–601.

[18] T. Söderström, *Errors-in-Variables Methods in System Identification*. Springer, 2018.

[19] G. Golub, A. Hoffman, and G. Stewart, "A generalization of the Eckart–Young–Mirsky matrix approximation theorem," *Linear Algebra Appl.*, vol. 88/89, pp. 317–327, 1987.

[20] G. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, pp. 211–218, 1936.

[21] I. Markovsky and K. Usevich, "Software for weighted structured low-rank approximation," *J. Comp. Appl. Math.*, vol. 256, pp. 278–292, 2014.

[22] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: Theory, implementation, applications*. Boston: Kluwer, 1996.