A missing data approach to data-driven filtering and control

Ivan Markovsky

***Abstract*—In filtering, control, and other mathematical engineering areas it is common to use a model-based approach, which splits the problem into two steps: 1) model identification and 2) model-based design. Despite its success, the model-based approach has the shortcoming that the design objective is not taken into account at the identification step, *i.e.*, the model is not optimized for its intended use. This paper proposes an approach for data-driven filtering and control that combines the identification and the model-based design into one joint problem. The signal of interest is modeled as a missing part of a trajectory of the data generating system. Subsequently, the missing data estimation problem is reformulated as a mosaic-Hankel structured matrix low-rank approximation/completion problem. A local optimization method, based on the variable projections principle, is then used for its numerical solution. The missing data estimation approach and the solution method proposed are illustrated on filtering and smoothing examples.**

***Index Terms*—Behavioral approach, System identification, Data-driven filtering, Structured low-rank approximation, Missing data.**

I. Introduction and context

The context of this work is an alternative paradigm, called data-driven, to the classical model-based paradigm. After introducing the data-driven paradigm, we describe informally the main contribution of this paper: posing and solving data-driven filtering and control problems as missing data estimation.

A. Model-based vs data-driven filtering and control

State-of-the-art signal processing and control methods are model-based. First, a model class is selected using prior knowledge and observed data. Then, model parameters are estimated using the data. Finally, the filtering/control task is solved using the identified model and the design specification. The model-based approach splits the original problem into

1) model identification [1], [2] and
2) model-based design,

which are solved independently.

There is a much work done separately on identification and model-based design, but relatively little work on their interplay in solving the overall problem. The cliche "all models are wrong but some are useful" is true when the model-based methods are applied in practice, where there is no "true" model in the model class. The question occurs "*what is the best model for the problem at hand?*" The identification literature answers instead questions about the closeness of the identified model to a true model and does not take into account the subsequent usage of the model for model-based design, *e.g.*, noise filtering, prediction, and control.

The issue of developing identification methods aimed at their intended usage is considered in an area of research known as "identification for control", see, *e.g.*, [3]. The identified model is tuned for maximum performance of the closed-loop system, *i.e.*, the identification criterion is linked with the control objective. The interplay between identification and control is central also in adaptive control, where the modeling and control tasks are solved simultaneously, in real-time. Both identification for control and adaptive control, however, consider model-based methods.

An alternative to the model-based approach is to solve the filtering or control problem directly without first identifying a model, see Figure 1. From applications' point of view, this what we call data-driven approach is closer to the real-life problem than the model-based approach. Indeed, in practice model parameters are rarely given, but data may often be observed. From the theoretical point of view, a data-driven approach opens up the possibility for a new class of solution methods and algorithms not based on an explicit model representation of the data generating process.

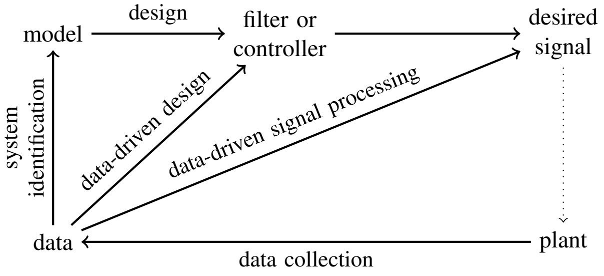

Fig. 1. Data-driven methods bypass the model identification step. Such method map plant data to controller/filter or directly to the desired signal.

B. Literature review

Data-driven control, also known as model-free control, has its roots in classical heuristics for PID controller tuning such as the Ziegler–Nichols method [4]. Rigorous data-driven methods, however, appeared only in the late 90's [5], [6], [7], [8]. Since then data-driven control has gained a lot of interest as evident by the large number of publications.

Although the particular problems considered range from LQG to fuzzy control, the corresponding methods developed can be classified into three main approaches:

- *Subspace-type data-driven methods* are proposed for solution of $\mathscr{H}_2/\mathscr{H}_\infty$ control problems in [8], [9], [10], [11]. The signal of interest is constrained to belong to a subspace computed from the measured data only. Viewed abstractly, the subspace is a model for the signal, although it is not parameterized in a familiar transfer function or state-space form.
- An adaptive approach, known as *controller unfalsification*, is developed in [6], [12], [13]. In this approach, the controller is viewed as an exclusion rule [14] and the main idea is to reject (falsify) controllers using previously collected experimental data from the plant.
- In *iterative feedback tuning* the controller parameters are optimized by a gradient type method minimizing the control objective, which depends on measured data only [7], [15].

For more details about the methods and an extensive list of references, we refer the reader to the recent overview paper [16].

The missing data approach proposed in this paper differs significantly from the existing approaches for data-driven control. The emphasis in this work is on the combination of the system identification and design objectives in one joint problem and posing it as a mosaic-Hankel structured low-rank matrix approximation and completion problem for which existing methods exist.

C. Missing data approach to data-driven filtering/control

The classical motivation for missing data in signal processing and control problems is sensor failures, where measurements are *accidentally* corrupted. More recently, missing data estimation is used for compressive sensing, where measurements are *intentionally* skipped. The main contribution of this paper, is in using missing

I. Markovsky is with the Department ELEC, Vrije Universiteit Brussel (VUB), 1050 Brussels, Belgium (e-mail: ivan.markovsky@vub.ac.be)

data for solving data-driven estimation and control problems, *i.e.*, the missing data represents the object that we *aim to find* on the first place. Examples are initial state state estimation, prediction, smoothing, partial realization, and optimal tracking control.

We pose the data-driven filtering/control problem as the problem of finding a missing part of a trajectory of a linear time-invariant (LTI) system, where other parts of the trajectory are given and are exact or are approximated. Once the problem is formulated as a missing values estimation, it is shown to be equivalent to a element-wise weighted mosaic-Hankel structured low-rank matrix approximation and completion (WSLRAC) problem. The latter problem is well researched in linear algebra and optimization. Theoretical results and effective solution methods with provable properties exist. We use a method based on the variable projections principle [17]. This choice is motivated by the existence of efficient algorithms that converge globally to a local minimum point with super linear convergence rate, have linear computational cost in the number of data points, and are implemented in a readily available software. The reformulation of the data-driven filtering/control problem as a WSLRAC problem is stated in Section V, Proposition 4, and is our key result.

The paper is organized as follows. After summarizing in Section II the basic concepts and notation used, we present in Section III the main idea: posing data-driven problems as missing data estimation. Specific examples are shown in Section IV in order to illustrate the idea. In Section V we describe the solution approach. First, we establish the equivalence of the data-driven problem and a WSLRA problem. Then, we describe a local optimization method based on the variable projections principle. Finally, we establish properties of the method in the errors-in-variables setting. In Section VI we show simulation examples that illustrate the theoretical properties and compare the data-driven method with classical model-based methods. Perspective for future work are given in Section VII.

II. Notation and preliminaries

Consider the class $\mathscr{L}^q$ of finite-dimensional, q-variate, discrete-time, LTI systems. A trajectory w of such a system is a vector valued sequence $w(1), w(2), \dots$, where $w(t) \in \mathbb{R}^q$. A system $\mathscr{B}$ is defined as the set of all its trajectories. The notation $w \in \mathscr{B}$ is a short-hand for "w is a trajectory of $\mathscr{B}$". $w_{\mathrm{p}} \wedge w_{\mathrm{f}}$ denotes the concatenation of the sequences w_{p} ("p" for "past") and w_{f} ("f" for "future").

Modulo a permutation Π of the variables, any trajectory $w \in \mathscr{B}$ has an input/output partition $\Pi w = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right]$, where u is an input (free variable), and y is an output (variable that is determined by the input, the system, and the initial conditions). In what follows, we assume that Π is the identity matrix I, *i.e.*, we assume that $w = (u, y) = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right]$. The number of inputs $\mathtt{m}$ and the number of outputs $\mathtt{p}$ of a system $\mathscr{B}$ with $q = \mathtt{m} + \mathtt{p}$ variables are properties of the system and do not depend on the input/output partitioning.

Let σ be the shift operator

$$\sigma w(t) := w(t+1).$$

A system $\mathscr{B} \in \mathscr{L}^q$ admits a kernel representation

$$\ker\big(R(z)\big) := \{\, w \mid R_0 w + R_1 \sigma w + \cdots + R_\ell \sigma^\ell w = 0 \,\},$$

with parameter $R(z) = R_0 + R_1 z + \cdots + R_\ell z^\ell$, as well as an input/state/output representation

$$\mathscr{B}_{\mathrm{i/s/o}}(A,B,C,D) := \{\, w = (u,y) \mid \text{there is } x, \text{ such that} \\ \sigma x = Ax + Bu \text{ and } y = Cx + Du \,\},$$

with parameters $A \in \mathbb{R}^{\mathtt{n} \times \mathtt{n}}$, $B \in \mathbb{R}^{\mathtt{n} \times \mathtt{m}}$, $C \in \mathbb{R}^{\mathtt{p} \times \mathtt{n}}$, and $D \in \mathbb{R}^{\mathtt{p} \times \mathtt{m}}$. The state dimension $\mathtt{n}$ is called the order of the state space representation. An input/state/output representation $\mathscr{B}_{\mathrm{i/s/o}}(A,B,C,D)$ is minimal if its order is as small as possible. This smallest possible order $\mathtt{n}(\mathscr{B})$ is invariant of the representation and is called the order of the system. Another invariant that is used in the paper is the lag $\ell(\mathscr{B})$ of $\mathscr{B}$. It is defined as the observability index of an input/state/output representation $\mathscr{B}_{\mathrm{i/s/o}}(A,B,C,D)$ of $\mathscr{B}$, *i.e.*, the smallest integer ℓ, for which the observability matrix $\mathscr{O}_\ell(A,C) := \operatorname{col}(C, CA, \dots, CA^{\ell-1})$ has rank $\mathtt{n}(\mathscr{B})$. Alternatively, the lag of $\mathscr{B}$ is the smallest integer ℓ, for which there is a kernel representation $\mathscr{B} = \ker(R)$, with polynomial matrix R of degree ℓ.

A Hankel matrix with L block rows is denoted by

$$\mathscr{H}_L(w) := \begin{bmatrix} w(1) & w(2) & \cdots & w(T-L+1) \\ w(2) & w(3) & \cdots & w(T-L+2) \\ \vdots & \vdots & & \vdots \\ w(L) & w(L+1) & \cdots & w(T) \end{bmatrix}.$$

The signal w is called *persistently exciting* of order L if the Hankel matrix $\mathscr{H}_L(w)$ is of full row rank. The block matrix

$$\mathscr{H}_L(w^1, w^2) := \begin{bmatrix} \mathscr{H}_L(w^1) & \mathscr{H}_L(w^2) \end{bmatrix}$$

with Hankel blocks is called mosaic-Hankel matrix [18].

The block lower-triangular Toeplitz matrix with t block rows, composed of $H = \big(H(0), H(1), \dots\big)$ is denoted by

$$\mathscr{T}_t(H) := \begin{bmatrix} H(0) & 0 & \cdots & 0 \\ H(1) & H(0) & 0 & \vdots \\ \vdots & \ddots & \ddots & 0 \\ H(t-1) & \cdots & H(1) & H(0) \end{bmatrix}.$$

If H is the impulse response of $\mathscr{B}_{\mathrm{i/s/o}}(A,B,C,D)$, we have that

$$H(0) = D, \qquad H(\tau) = CA^{\tau-1}B, \quad \text{for } \tau = 1, 2, \dots, t-1.$$

Using the notation $\mathscr{O}$ for the observability matrix and the notation $\mathscr{T}$ for the lower triangular Toeplitz matrix, we can express the condition that w is a trajectory of $\mathscr{B}$ as a linear system of equations

$$w = (u,y) \in \mathscr{B} \quad \iff \quad \text{there is } x_{\mathrm{ini}} \in \mathbb{R}^{\mathtt{n}}, \\ \text{such that } y = \mathscr{O}_t(A,C) x_{\mathrm{ini}} + \mathscr{T}_t(H) u.$$

Note that with some abuse of notation, we use y for both the signal $\big(y(1), \dots, y(t)\big)$ and the vector $\begin{bmatrix} y^\top(1) & \cdots & y^\top(t) \end{bmatrix}^\top$.

A candidate model $\widehat{\mathscr{B}}$ for a time series w_{d} ("d" for "data") is *unfalsified* if $w_{\mathrm{d}} \in \widehat{\mathscr{B}}$. The *most powerful unfalsified model* for the data w_{d} in the model class $\mathscr{L}^q$ is defined as

$$\mathscr{B}_{\mathrm{mpum}}(w_{\mathrm{d}}) := \arg \underbrace{\min_{\widehat{\mathscr{B}} \in \mathscr{L}^q} \ell(\widehat{\mathscr{B}})}_{\text{most powerful}} \quad \text{subject to} \quad \underbrace{w_{\mathrm{d}} \in \widehat{\mathscr{B}}}_{\text{unfalsified model}} . \tag{1}$$

The system $\mathscr{B}_{\mathrm{mpum}}(w_{\mathrm{d}})$ is the least complicated linear exact time-invariant model for the data w_{d}. The model complexity is measured by its lag $\ell(\mathscr{B})$. The subclass of $\mathscr{L}^q$ with at most $\mathtt{m}$ inputs and lag at most ℓ (models of bounded complexity) is denoted by $\mathscr{L}_{\mathtt{m},\ell}$. $\|\cdot\|$ is the Euclidean norm.

III. The missing data approach

In data-driven problems, instead of the plant $\mathscr{B}$, given is a trajectory $w_{\mathrm{d}} = \big(w_{\mathrm{d}}(1), \dots, w_{\mathrm{d}}(T_{\mathrm{d}})\big)$ of $\mathscr{B}$. Under controllability and persistency of excitation assumptions, w_{d} completely specifies $\mathscr{B}$.

Lemma 1 (Identifiability conditions [19])**.** *Let $w_d = (u_d, y_d)$ be an exact trajectory of a controllable LTI system $\mathscr{B}$ and let the input u_d be persistently exciting of order $\mathtt{n}(\mathscr{B}) + \ell(\mathscr{B}) + 1$. Then, the most powerful unfalsified model of w_d coincides with the data generating system,* i.e., *$\mathscr{B}_{mpum}(w_d) = \mathscr{B}$.*

In the classical model-based approach, w_d is refered to as the "identification data" and the first step towards solving filtering and control problems is to compute an explicit representation of the data generating system, *e.g.*, a state space representation $\mathscr{B}_{\mathrm{i/s/o}}(A,B,C,D) = \mathscr{B}_{\mathrm{mpum}}(w_\mathrm{d})$. In the data-driven approach, we use the data w_d directly in an optimization problem with combined objective.

In order to develop a general approach that unifies various problems, we consider a generic trajectory $w = \big(w(1),\dots,w(T)\big)$ of the system $\mathscr{B}$, partition the variables into inputs u and outputs y, and split the time axis into "past"—the first T_p samples—and "future"—the remaining T_f samples:

	past	future
input	u_p	u_f
output	y_p	y_f

When the past horizon T_p is sufficiently long, *i.e.*, $T_\mathrm{p} \geq \ell(\mathscr{B})$, w_p completely specifies the initial conditions for the future trajectory w_f.

Lemma 2 (Initial condition w_p, [11])**.** *Let $\mathscr{B}_{i/s/o}(A,B,C,D)$ be a minimal input/state/output representation and let H be the impulse response of $\mathscr{B} \in \mathscr{L}^q$. Then for all $w_p \in (\mathbb{R}^q)^{T_p}$, with $T_p \geq \ell(\mathscr{B})$,*

$$w_p \wedge (u_f, y_f) \in \mathscr{B} \quad \Longrightarrow \quad \text{there is unique } x_{ini} \in \mathbb{R}^{\mathtt{n}(\mathscr{B})}, \\ \text{such that } y_f = \mathscr{O}_{T_f}(A,C)x_{ini} + \mathscr{T}_{T_f}(H)u_f.$$

The signal we aim to compute in the data processing problem is an unknown part of a trajectory w of $\mathscr{B}$. Therefore, the trajectory w includes missing elements. In addition to the missing values, w has exact and inexact (noisy) elements. The exact elements correspond to specification of the to-be-found signal, *e.g.*, the impulse response is specified by zero initial conditions and pulse input. The inexact elements represent part of the signal that has to be approximated due to, *e.g.*, additive measurement noise.

The general approach of representing the to-be computed signal as missing data in a trajectory with missing, exact, and noisy elements is illustrated next by the classical examples of state estimation, noise filtering/smoothing, simulation, and tracking control.

IV. Examples

First, we state the classical model-based problems. Then, we state the corresponding data-driven problems, where the system $\mathscr{B}$ is implicitly specified by data w_d. Finally, the problems are formulated as missing data estimation in terms of the a trajectory w. Each of the elements u_p, y_p, u_f, and y_f of w is exact, inexact, or missing depending on the particular problem.

A. State estimation and Kalman smoothing

The classical model-based state estimation problem is defined as follows: given an LTI system $\mathscr{B}$ and a trajectory w_f,

$$\text{find } w_\mathrm{p}, \text{ such that } w = w_\mathrm{p} \wedge w_\mathrm{f} \in \mathscr{B}. \tag{2}$$

The aim of (2) is to estimate of the first T_p samples of a trajectory w, with the other samples w_f known exactly.

If w_f is not a trajectory of $\mathscr{B}$, the model-based state estimation problem becomes the famous Kalman smoothing problem. The classical Kalman smoother [20] assumes that the input u_d is exact, in which case the approximation problem is

$$\begin{aligned} &\text{minimize} \quad \text{over } \widehat{w}_\mathrm{p} \text{ and } \widehat{y}_\mathrm{f} \quad \|y_\mathrm{f} - \widehat{y}_\mathrm{f}\| \\ &\text{subject to} \quad \widehat{w}_\mathrm{p} \wedge (u_\mathrm{f}, \widehat{y}_\mathrm{f}) \in \mathscr{B}. \end{aligned} \tag{3}$$

As a byproduct of computing the initial conditions estimate $\widehat{w}_\mathrm{p}$, (3) determines an approximation of the output $\widehat{y}_\mathrm{f}$ (the smoothed output). The signal $\widehat{y}_\mathrm{f}$ is the best estimate of the noisy output y_f, given the model $\mathscr{B}$. Problem (3) is also a missing data estimation problem, however, the output y_f is approximated rather than fitted exactly.

When both u_f and y_f are inexact, the smoothing problem is

$$\begin{aligned} &\text{minimize} \quad \text{over } \widehat{w}_\mathrm{p} \text{ and } \widehat{w}_\mathrm{f} \quad \|w_\mathrm{f} - \widehat{w}_\mathrm{f}\| \\ &\text{subject to} \quad \widehat{w}_\mathrm{p} \wedge \widehat{w}_\mathrm{f} \in \mathscr{B}. \end{aligned} \tag{4}$$

and is refered to as the errors-in-variables (EIV) Kalman smoother. The solution of (4) is given by a modification of the ordinary Kalman smoother, see [21]. The resulting algorithm employs a Riccati-type recursion and has linear computational complexity in the number of samples T_f.

The data-driven version of the state estimation problem is: given trajectories w_d and w_f of an LTI system $\mathscr{B}$,

$$\text{find } w_\mathrm{p}, \text{ such that } w = w_\mathrm{p} \wedge w_\mathrm{f} \in \mathscr{B}_{\mathrm{mpum}}(w_\mathrm{d}). \tag{5}$$

Although the data-driven problem formulation involves the most powerful unfalsified model of the data, solution methods need not identify explicitly a representation of $\mathscr{B}_{\mathrm{mpum}}(w_\mathrm{d})$ in order to find the quantity of interest w_p.

When w_d is inexact, prior knowledge about the model is needed. Often, it is the lag ℓ of the data generating system, which determines the model class $\mathscr{L}_{\mathtt{m},\ell}$, to which the system belongs. Then the data-driven versions of the state estimation problems (3) and (4) are

$$\begin{aligned} &\text{minimize over } \widehat{w}_\mathrm{d} \text{ and } \widehat{y}_\mathrm{f} \quad \underbrace{\|y_\mathrm{f} - \widehat{y}_\mathrm{f}\|_2^2}_{\text{estimation error}} + \underbrace{\|w_\mathrm{d} - \widehat{w}_\mathrm{d}\|_2^2}_{\text{identification error}} \\ &\text{subject to} \quad (u_\mathrm{f}, \widehat{y}_\mathrm{f}) \in \mathscr{B}_{\mathrm{mpum}}(\widehat{w}_\mathrm{d}) \in \mathscr{L}_{\mathtt{m},\ell} \end{aligned} \tag{6}$$

and

$$\begin{aligned} &\text{minimize over } \widehat{w}_\mathrm{d} \text{ and } \widehat{w} \quad \underbrace{\|w_\mathrm{f} - \widehat{w}_\mathrm{f}\|_2^2}_{\text{estimation error}} + \underbrace{\|w_\mathrm{d} - \widehat{w}_\mathrm{d}\|_2^2}_{\text{identification error}} \\ &\text{subject to} \quad \widehat{w} \in \mathscr{B}_{\mathrm{mpum}}(\widehat{w}_\mathrm{d}) \in \mathscr{L}_{\mathtt{m},\ell}, \end{aligned} \tag{7}$$

respectively.

The classical approach for state estimation involves the two steps:

1) *identification:* given w_d and ℓ, compute a representation of $\mathscr{B} = \mathscr{B}_{\mathrm{mpum}}(\widehat{w}_\mathrm{d})$, where $\widehat{w}_\mathrm{d} = w_\mathrm{d}$, if w_d is exact, or compute a solution $\widehat{w}_\mathrm{d}$ of the optimization problem

$$\begin{aligned} &\text{minimize} \quad \text{over } \widehat{w}_\mathrm{d} \quad \|w_\mathrm{d} - \widehat{w}_\mathrm{d}\| \\ &\text{subject to} \quad \mathscr{B}_{\mathrm{mpum}}(\widehat{w}_\mathrm{d}) \in \mathscr{L}_{\mathtt{m},\ell} \end{aligned} \tag{8}$$

   if w_d is inexact;

2) *model-based design:* solve (2), (3), or (4), using the representation of $\mathscr{B}$ computed on step 1.

Note that the optimization criterion of the data-driven problem (7) involves a mixture of the identification and filtering/control errors, while (8) is agnostic to the filter/control design objective.

B. Other examples

Other examples that fit into the generic approach for data-driven filtering/control, presented in Section III are simulation, partial realization, and output tracking.

- *Simulation:* Given initial conditions w_p and input u_f, the objective is to find the corresponding output y_f of the system, *i.e.*,

$$\text{find } y_\mathrm{f}, \text{ such that } w_\mathrm{p} \wedge (u_\mathrm{f}, y_\mathrm{f}) \in \mathscr{B}. \tag{9}$$

- *Noisy partial realization [22], [23]:* given the first T samples $H(1),\dots,H(T)$ of an impulse response, the objective of the partial realization problem is to find the remaining samples $H(T+1), H(T+2),\dots$ of the impulse response. Partial realization is a fundamental problem in system theory and is the basis for the class of subspace identification methods.

- *Output tracking:* given initial conditions w_{p}, and an output y_{f}, the objective is to find a control input u_{f}, such that

$$\begin{array}{ll}\text{minimize} & \text{over } \widehat{u}_{\mathrm{f}},\ \widehat{y}_{\mathrm{f}} \quad \|y_{\mathrm{f}} - \widehat{y}_{\mathrm{f}}\| \\ \text{subject to} & w_{\mathrm{ini}} \wedge (\widehat{u}_{\mathrm{f}}, \widehat{y}_{\mathrm{f}}) \in \mathscr{B}.\end{array} \tag{10}$$

The signal u_{f} is the *open-loop* optimal control signal.

Table I gives a summary of the examples reviewed.

TABLE I
SUMMARY OF THE EXAMPLES.
LEGEND: ? — MISSING, E — EXACT, N — NOISY/INEXACT.

example	reference	u_{p}	y_{p}	u_{f}	y_{f}
simulation	(9)	E	E	E	?
partial realization	[22]	E	E	E	E/?
state estimation	(2)	?	?	E	E
classical Kalman smoothing	(3)	?	?	E	N
EIV Kalman smoothing	(4)	?	?	N	N
noisy realization	[23]	E	E	E	N/?
output tracking	(10)	E	E	?	N

V. POSING THE PROBLEM AS STRUCTURED LOW-RANK APPROXIMATION/COMPLETION

First, we show the link between data-driven filtering/control and mosaic-Hankel WSLRAC. Then, we present a method based on local optimization and the variable projections principle [17].

A. Link to weighted mosaic-Hankel low-rank approximation

The data-driven problems, considered in Section IV, aim to minimize the "size" of the error signal $e := w - \widehat{w}$, where w contains given data (exact or noisy) as well as missing values and $\widehat{w}$ is a trajectory of the system. We encode the information about exact, noisy, and missing data by the weights $v_i(t) \geq 0$ of the semi-norm

$$\|e\|_v := \sqrt{\textstyle\sum_{t=1}^{T} \sum_{i=1}^{q} \ v_i(t) e_i^2(t)}.$$

TABLE II
THE INFORMATION ABOUT EXACT, NOISY, AND MISSING DATA ELEMENTS $w_i(t)$ IS ENCODED IN THE WEIGHTS $v_i(t)$ OF THE SEMI-NORM $\|\cdot\|_v$.

weight	used	to	by
$v_i(t) = \infty$	if $w_i(t)$ is exact	interpolate $w_i(t)$	$e_i(t) = 0$
$v_i(t) \in (0, \infty)$	if $w_i(t)$ is noisy	approximate $w_i(t)$	min $\|e_i(t)\|$
$v_i(t) = 0$	if $w_i(t)$ is missing	fill in $w_i(t)$	$\widehat{w} \in \widehat{\mathscr{B}}$

With this notation, the examples of data-driven problems, shown in Table I, become special cases of the following generic problem

$$\begin{array}{ll}\text{minimize} & \text{over } \widehat{w}_{\mathrm{d}},\ \widehat{w} \ \|w_{\mathrm{d}} - \widehat{w}_{\mathrm{d}}\|_2^2 + \|w - \widehat{w}\|_v^2 \\ \text{subject to} & \widehat{w} \in \mathscr{B}_{\mathrm{mpum}}(\widehat{w}_{\mathrm{d}}) \in \mathscr{L}_{\mathtt{m},\ell},\end{array} \tag{11}$$

for a suitable choice of the trajectory w and the weights v.

In order to solve (11), we use the equivalence of trajectories of an LTI system with bounded complexity and rank deficiency of a mosaic-Hankel matrix constructed from these trajectories.

Lemma 3. *Let* $\mathtt{p}$ *and* ℓ *be, respectively, the number of outputs and the lag of an LTI system* $\mathscr{B}$. *Then,*

$$w^1, w^2 \in \mathscr{B} \quad \iff \quad rank\big(\mathscr{H}_{\ell+1}(w^1, w^2)\big) \leq q\ell + \mathtt{m}.$$

Proof. Let $\mathscr{B} = \ker\big(R(z)\big)$ be a kernel representation of the system.

$$\begin{aligned} w^i \in \mathscr{B} \in \mathscr{L}_{\mathtt{m},\ell} \quad &\iff \\ &\underbrace{\begin{bmatrix} R_0 & R_1 & \cdots & R_\ell \end{bmatrix}}_{R} \mathscr{H}_{\ell+1}(w^i) = 0, \ \text{for } i = 1, 2. \end{aligned} \tag{12}$$

The $\mathtt{p} \times q(\ell+1)$ matrix R is full row-rank [24]. Then

$$\begin{aligned} &\begin{cases} R \in \mathbb{R}^{\mathtt{p} \times q(\ell+1)} \ \text{full row rank} \\ R \begin{bmatrix} \mathscr{H}_{\ell+1}(w^1) & \mathscr{H}_{\ell+1}(w^2) \end{bmatrix} = 0 \end{cases} \\ &\qquad\qquad \iff \quad \text{rank}\big(\mathscr{H}_{\ell+1}(w^1, w^2)\big) \leq q\ell + \mathtt{m}. \end{aligned}$$

□

Using Lemma 3, we obtain an equivalent mosaic-Hankel WSLRAC problem to the data-driven problem (11).

Proposition 4. *Problem (11) is equivalent to the following mosaic-Hankel WSLRAC problem*

$$\begin{array}{ll}\textit{minimize} & \textit{over } \widehat{w}_d,\ \widehat{w} \quad \|w_d - \widehat{w}_d\|_2^2 + \|w - \widehat{w}\|_v^2 \\ \textit{subject to} & \textit{rank}\big(\mathscr{H}_{\ell+1}(\widehat{w}_{\mathrm{d}}, \widehat{w})\big) \leq q\ell + \mathtt{m}.\end{array} \tag{13}$$

Proof. We need to show that the constraint of (11) is equivalent to the constraint of (13). The constraint

$$\widehat{w} \in \mathscr{B}_{\mathrm{mpum}}(\widehat{w}_{\mathrm{d}}) \in \mathscr{L}_{\mathtt{m},\ell}$$

of (11) is equivalent to the existence of an exact model $\widehat{\mathscr{B}} \in \mathscr{L}_{\mathtt{m},\ell}$ for both $\widehat{w}$ and $\widehat{w}_{\mathrm{d}}$. By Lemma 3, $\widehat{w}, \widehat{w}_{\mathrm{d}} \in \widehat{\mathscr{B}} \in \mathscr{L}_{\mathtt{m},\ell}$ is equivalent to

$$\text{rank}\big(\mathscr{H}_{\ell+1}(\widehat{w}_{\mathrm{d}}, \widehat{w})\big) \leq q\ell + \mathtt{m},$$

which is the constraint of (13). □

Problem (13) is a nonconvex optimization problem. It can be solved by convex relaxation, using the nuclear norm heuristic, subspace methods, and local optimization methods. Next, we describe a local optimization method, based on the variable projections principle.

B. Solution method based on the variable projections

The rank constraint in (13) is expressed as a condition on the dimension of the left kernel of the matrix $\mathscr{H}_{\ell+1}(\widehat{w}_{\mathrm{d}}, \widehat{w})$

$$\begin{aligned} \text{rank}\big(\mathscr{H}_{\ell+1}(\widehat{w}_{\mathrm{d}}, \widehat{w})\big) \leq q\ell + \mathtt{m} \quad &\iff \\ &\text{there is full row rank } R \in \mathbb{R}^{\mathtt{p} \times q(\ell+1)}, \\ &\text{such that } R\mathscr{H}_{\ell+1}(\widehat{w}_{\mathrm{d}}, \widehat{w}) = 0. \end{aligned} \tag{14}$$

Then, (13) is equivalent to

$$\begin{array}{l}\text{minimize over } \widehat{w}_{\mathrm{d}},\ \widehat{w},\ R \in \mathbb{R}^{\mathtt{p} \times q(\ell+1)} \ \|w_{\mathrm{d}} - \widehat{w}_{\mathrm{d}}\|_2^2 + \|w - \widehat{w}\|_v^2 \\ \text{subject to} \quad R\mathscr{H}_{\ell+1}(\widehat{w}_{\mathrm{d}}, \widehat{w}) = 0 \quad \text{and} \quad R \text{ is full row rank.}\end{array} \tag{15}$$

The variables $\widehat{w}_{\mathrm{d}}$ and $\widehat{w}$ can be eliminated by representing (15) as a double minimization problem:

$$\text{minimize} \quad \text{over full row rank } R \in \mathbb{R}^{\mathtt{p} \times q(\ell+1)} \quad M(R), \tag{16}$$

where

$$\begin{aligned} M(R) := \min_{\widehat{w}_{\mathrm{d}}, \widehat{w}} \quad &\|w_{\mathrm{d}} - \widehat{w}_{\mathrm{d}}\|_2^2 + \|w - \widehat{w}\|_v^2 \\ \text{subject to} \quad &R\mathscr{H}_{\ell+1}(\widehat{w}_{\mathrm{d}}, \widehat{w}) = 0. \end{aligned} \tag{17}$$

Solution of (17), *i.e.*, evaluation of $M(R)$ for given R, is refered to as the *inner minimization*. Solution of (16), *i.e.*, optimization of M over R, is referred to as the *outer minimization*.

The inner minimization problem (17) is a generalized linear least squares problem [25] and admits an analytic solution. In the case of no missing values, M can be evaluated with a linear cost in the number of data points $T_{\mathrm{d}} + T_{\mathrm{f}}$. Fast algorithms for missing data estimation is a topic of current research.

The advantage of reformulating (15) as (16) is the elimination of the optimization variables $\widehat{w}_{\mathrm{d}}$ and $\widehat{w}$. In applications of filtering and control, $\widehat{w}_{\mathrm{d}}$ and $\widehat{w}$ are high dimensional and R is small dimensional. Therefore, the elimination of $\widehat{w}_{\mathrm{d}}$ and $\widehat{w}$ leads to a big reduction in the number of the optimization variables. The approach for solving (15) described above is similar to the variable projection method [17] for separable unconstrained non-linear least squares minimization.

In (16), the cost function M is minimized over the set of full row rank matrices R. Since $M(R)$ depends only on the space spanned by the rows of R, *i.e.*, $M(R) = M(UR)$, for all nonsingular $U \in \mathbb{R}^{\mathtt{p}\times\mathtt{p}}$, (16) is a minimization problem on a Grassmann manifold (the $q\ell+\mathtt{m}$-dimensional subspaces of $\mathbb{R}^{q(\ell+1)}$) [26], [27]. In [27] the optimization over a manifold problem is reduced to a classical unconstrained optimization over an Euclidean space. Subsequently, standard local optimization methods such as the Levenberg-Marquardt method [28] can be used for its solution. The resulting variable projections method for WSLRAC inherits the convergence properties of the standard local optimization method being used (global convergence with super-linear convergence rate in the case of the Levenberg-Marquardt method). By default an initial approximation is computed by interpolating the missing values and doing low-rank approximation (singular value decomposition) of the resulting matrix, *i.e.*, we ignore the structure constraint and replace the weighted norm by the 2-norm.

A software package for solving (13), based on the variable projections approach, is developed in [29], [30] and is used in Section VI. The function `ident` solves problem (11) and the function `misfit` solves the inner minimization problem (17).

C. Properties of the estimators in the errors-in-variables setting

In the errors-in-variables setting [31] the data w_d is obtained as

$$w_\mathrm{d} = \overline{w}_\mathrm{d} + \widetilde{w}_\mathrm{d}, \tag{18}$$

where $\overline{w}_\mathrm{d}$, called true value of w_d, is a trajectory of a model $\bar{\mathscr{B}} \in \mathscr{L}_{\mathtt{m},\ell}$ and $\widetilde{w}_\mathrm{d}$, called measurement noise, is a realization of a zero mean white Gaussian process. Then, the statistically optimal choice of the weights v is $v_i(t) = 1/\sigma_i(t)$, where $\sigma_i(t)$ is the standard deviation of the measurement noise on $w_{\mathrm{d},i}(t)$. In this case, minimization of the criterion $\|w_\mathrm{d} - \widehat{w}\|_v$, subject to the constraint $\widehat{w} \in \mathscr{B}$ leads to the *maximum-likelihood estimator* [32]. By standard results [33], it follows that the estimator is consistent and efficient.

Note 5 (Stochastic interpretation of zero and infinite weights). If $w_i(t)$ is known exactly, the noise standard deviation is zero and the corresponding weight in the cost function is infinite. Infinite weight imposes an implicit equality constraint $w_i(t) = \widehat{w}_i(t)$, *i.e.*, the approximation $\widehat{w}$ agrees with the data w for the variable i at time t. If $w_i(t)$ is missing, the noise standard deviation is infinite and the corresponding weight is zero. Zero weight excludes the element $w_i(t)$ from the cost function. The approximation $\widehat{w}_i(t)$ is then determined solely from the constraint $\widehat{w} \in \widehat{\mathscr{B}}$.

Next, we compare the solution $\widehat{w}$ of the data-driven approach (11) with the solution $\widehat{w}'$ of the classical model-based approach:

$$w_\mathrm{d} \xrightarrow{\substack{\text{system}\\ \text{identification}}} \widehat{\mathscr{B}}' \xrightarrow{\substack{\text{model-based}\\ \text{design}}} \widehat{w}'.$$

In two cases—exact identification and exact design—the solutions coincide. In other cases, the data-driven approach, being statistically optimal, gives more accurate estimates than the classical approach.

Consider first, the exact data case. Under the assumptions of Lemma 1, the identification error $\|w_\mathrm{d} - \widehat{w}\|_2 = 0$, and by Lemma 1, $\widehat{\mathscr{B}} = \mathrm{mpum}(w_\mathrm{d}) = \bar{\mathscr{B}}$. Therefore, the data-driven problem (11) is equivalent to the model-based design problem. Consider, next exact design case, *i.e.*, $\|w - \widehat{w}\|_v = 0$, for all $\widehat{w}$, for example data-driven simulation (9). In this case, the cost function of the data-driven problem (11) is equal to the cost function of the identification problem (8). Since the constraint $\widehat{w}_\mathrm{d} \in \widehat{\mathscr{B}} \in \mathscr{L}_{\mathtt{m},\ell}$ is the same in (11) and (8), the two problems are equivalent and therefore $\widehat{w} = \widehat{w}'$.

An intuitive explanation why the data-driven approach is superior to the model-based one can be given in the errors-in-variables smoothing problem (7). The data-driven problem (11) uses as identification data two trajectories—w_d and w_f—while the classical method uses only one trajectory w_d. Then, by [34, Section V], $\widehat{w}$ is statistically more accurate estimate than $\widehat{w}'$. This statement is illustrated in the next section on simulation examples.

VI. NUMERICAL EXPERIMENTS

In this section, we show simulation examples of the matrix completion approach for data-driven signal processing, implemented in the variable projections method, described in Section V. The problems considered are state estimation from step response data and Kalman smoothing in the errors-in-variables setting. The data generating system $\mathscr{B}$ is LTI

$$\mathscr{B} = \{\, (u,y) \mid u - \sigma u + \sigma^2 u = 0.81y - 1.456\sigma y + \sigma^2 y \,\}.$$

First, we solve the problems, assuming that the model is given, however, the computation of the quantity of interest is done by missing data estimation with the function `misfit`. The result obtained is compared with the results obtained by classical model-based methods (Kalman smoother and matrix exponential). Since missing data estimation solves the same problem as the classical model-based method, the aim of the comparison is an empirical confirmation that the solutions obtained are the same. Then, we solve the problem use data obtained in the EIV setup (18) rather than the model $\mathscr{B}$.

The task is to find the smoothed signal $\widehat{w}_\mathrm{f}$ for a given noisy trajectory $w_\mathrm{f} = \overline{w}_\mathrm{f} + \widetilde{w}_\mathrm{f}$, obtained as a true value $\overline{w}_\mathrm{f} \in \mathscr{B}$ plus noise, where the noise $\widetilde{w}_\mathrm{f}$ is a realization of a zero mean, white, Gaussian process. The true trajectory $\overline{w}_\mathrm{f}$ consists of the first T_f samples of the step response of $\mathscr{B}$. The estimation methods are compared in terms of the relative approximation errors $e := (\|\overline{w}_\mathrm{f} - \widehat{w}_\mathrm{f}\|)/\|\overline{w}_\mathrm{f}\|$ with respect to the true trajectory $\overline{w}_\mathrm{f}$. With known true data generating system $\mathscr{B}$, the optimal least-squares estimator is the EIV Kalman smoother (4). The "classical" method for solving this problem is using a state-space representation $\mathscr{B}_{\mathrm{i/s/o}}(A,B,C,D)$ of $\mathscr{B}$. (4) is equivalent to the following linear least squares problem:

$$\min_{\widehat{u},x_\mathrm{ini}} \left\| \begin{bmatrix} u \\ y \end{bmatrix} - \begin{bmatrix} 0 & I \\ \mathscr{O}_T(A,C) & \mathscr{T}_T(H) \end{bmatrix} \begin{bmatrix} x_\mathrm{ini} \\ \widehat{u} \end{bmatrix} \right\|, \tag{19}$$

which admits a closed-form solution. The resulting method is implemented in a function `eiv_ks`. The result obtained with the classical state-space method is used to verify the solution obtained by the matrix completion/approximation method

$$\begin{aligned} &\text{minimize} \quad \text{over } \widehat{w}_\mathrm{p},\ \widehat{w}_\mathrm{f} \quad \|w_\mathrm{f} - \widehat{w}_\mathrm{f}\| \\ &\text{subject to} \quad R\mathscr{H}_3(\widehat{w}_\mathrm{p} \wedge \widehat{w}_\mathrm{f}) = 0, \end{aligned} \tag{20}$$

where R is a parameter of a kernel representation of the system (12). Problem (20) is solved by the function `misfit`.

Up to numerical errors of the computations, the relative errors for the `eiv_ks` and `misfit` functions are equal:

method	(19)	(20)
function	`misfit`	`eiv_ks`
error e	0.029358	0.029358

This is an empirical confirmation that matrix completion problem (20) is equivalent to the EIV Kalman smoothing problem (4).

Next, we solve the smoothing problem without knowledge of $\mathscr{B}$, using 1) classical approach of identification of a model $\widehat{\mathscr{B}}$ from the data w_d, followed by Kalman smoothing based on the identified model $\widehat{\mathscr{B}}$, 2) the data-driven approach (7), and 3) a subspace data-driven method `eiv_ks_dd`, derived using the methodology of [11].

The result obtained by the data-driven method (7) is worse than the one of the Kalman smoother using the true model but better than the one of the Kalman smoother using the identified model and the subspace data-driven method:

method	(7)	(8) + (20)	subspace
function	`ident`	`ident` + `eiv_ks`	`eiv_ks_dd`
error e	0.0310	0.0315	0.0412

The Kalman smoother based on the true model is statistically optimal in the setup of the simulation example. The superior performance of (7) over (8) + (20) is that the identified model in the classical approach (8) uses only the data w_d, while the data-driven method (7) uses as data both w_d and w_f. The fact that more data is used by the data-driven method in comparison with the classical method leads to more accurate results. Finally, the estimate obtained by the subspace data-driven method is suboptimal in the sense of the maximum-likelihood optimization criterion so that on average it can be expected to perform worse.

Simulation results for the other examples described in Section IV.B are presented in `http://slra.github.io/ddsp/demo.html`. The code needed to reproduce the results is also provided.

VII. Conclusions

Data-driven filtering and control deals with one joint problem formulation that involves both the data modeling objective and the design objective. We formulated the data-driven problem as estimation of a missing part in a trajectory of the (unknown) data-generating system. For example, state estimation, simulation, simulation, filtering/smoothing, partial realization, and output tracking control can be posed as missing data estimation problems. The missing data estimation problem is furthermore reformulated as an equivalent mosaic-Hankel WSLRAC problem. The implication of this fact is that existing methods, developed in the WSLRAC setting, are used for the numerical solution of the data-driven filtering/control problem. Simulation examples show the effectiveness of the local optimization methods based on the variable projection approach.

In order to make the missing data approach for data-driven filtering/control a practically feasible alternative to the model-based methods, fast algorithms with provable properties in the presence of measurement noise and disturbances need to be developed. This is a topic of current research. The main advantage of the approach presented in the paper is a common setting for posing different data-driven problems and solving them by different methods.

The methodology developed in this paper is currently limited to batch processing. In the context of control, the computed signal is applied to the plant in open loop. One way to use it in feedback control is to embed the batch computation (13) in an MPC-like scheme. Another way is to develop methods for recursive real-time solution of problem (13). Ensuring stability of the overall system is an open problem.

VIII. Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Programme (FP7/2007-2013) / ERC Grant agreement number 258581 "Structured low-rank approximation: Theory, algorithms, and applications".

References

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

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

[3] L. Ljung, "Identification for control: simple process models," in *Proceedings of the 41st IEEE Conference on Decision and Control, 2002*, vol. 4, Dec. 2002, pp. 4652–4657.

[4] J. Ziegler and N. Nichols, "Optimum settings for automatic controllers," *Trans. ASME*, vol. 64, pp. 759–768, 1942.

[5] J.-T. Chan, "Data-based synthesis of a multivariable linear-quadratic regulator," *Automatica*, vol. 32, pp. 403–407, 1996.

[6] M. Safonov, "Focusing on the knowable: Controller invalidation and learning," in *Control using logic-based switching*, A. Morse, Ed. Springer-Verlag, Berlin, 1996, pp. 224–233.

[7] H. Hjalmarsson, M. Gevers, S. Gunnarsson, and O. Lequin, "Iterative feedback tuning: theory and applications," *IEEE Control Systems Magazine*, vol. 18, pp. 26–41, 1998.

[8] W. Favoreel, "Subspace methods for identification and control of linear and bilinear systems," Ph.D. dissertation, ESAT, K.U.Leuven, 1999.

[9] G. Shi and R. Skelton, "Markov data-based LQG control," *J. of Dynamic Systems, Measurement, and Control*, vol. 122, pp. 551–559, 2000.

[10] B. Woodley, "Model free subspace based H_∞ control," Ph.D. dissertation, Stanford University, 2001.

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

[12] M. Safonov and T. Tsao, "The unfalsified control concept and learning," *IEEE Trans. Automat. Contr.*, vol. 42, no. 6, pp. 843–847, 1997.

[13] M. Safonov and F. Cabral, "Fitting controllers to data," *Control Lett.*, vol. 43, no. 4, pp. 299–308, 2001.

[14] J. Polderman and J. C. Willems, *Introduction to mathematical systems theory*. Springer-Verlag, 1998.

[15] R. Hildebrand, A. Lecchini, G. Solari, and M. Gevers, "Prefiltering in iterative feedback tuning: Optimization of the prefilter for accuracy," *IEEE Trans. Automat. Contr.*, vol. 49, pp. 1801–1806, 2004.

[16] Z.-S. Hou and Z. Wang, "From model-based control to data-driven control: Survey, classification and perspective," *Information Sciences*, vol. 235, pp. 3–35, 2013.

[17] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Institute of Physics, Inverse Problems*, vol. 19, pp. 1–26, 2003.

[18] G. Heinig, "Generalized inverses of Hankel and Toeplitz mosaic matrices," *Linear Algebra Appl.*, vol. 216, no. 0, pp. 43–59, Feb. 1995.

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

[20] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

[21] I. Markovsky and B. De Moor, "Linear dynamic filtering with noisy input and output," *Automatica*, vol. 41, no. 1, pp. 167–171, 2005.

[22] R. E. Kalman, "On partial realizations, transfer functions, and canonical forms," *Acta Polytechnica Scandinavica*, vol. 31, pp. 9–32, 1979.

[23] B. De Moor, "Total least squares for affinely structured matrices and the noisy realization problem," *IEEE Trans. Signal Proc.*, vol. 42, no. 11, pp. 3104–3113, 1994.

[24] I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor, *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*. SIAM, March 2006.

[25] C. Lawson and R. Hanson, *Solving Least Squares Problems*, ser. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1987.

[26] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press, 2008.

[27] K. Usevich and I. Markovsky, "Optimization on a Grassmann manifold with application to system identification," *Automatica*, vol. 50, pp. 1656–1662, 2014.

[28] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.*, vol. 11, pp. 431–441, 1963.

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

[30] I. Markovsky, "A software package for system identification in the behavioral setting," *Control Engineering Practice*, vol. 21, no. 10, pp. 1422–1436, 2013.

[31] T. Söderström, "Errors-in-variables methods in system identification," *Automatica*, vol. 43, pp. 939–958, 2007.

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

[33] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*, 2nd ed. Piscataway, NJ: IEEE Press, 2012.

[34] 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, 2015.