Approximate system identification with missing data

Ivan Markovsky

Abstract— **Linear time-invariant system identification is considered in the behavioral setting. Nonstandard features of the problem are specification of missing and exact variables and identification from multiple time series with different length. The problem is equivalent to mosaic Hankel structured low-rank approximation with element-wise weighted cost function. Zero/infinite weights are assigned to the missing/exact data points. The problem is in general nonconvex. A solution method based on local optimization is outlined and compared with alternative methods on simulation examples.**

In a stochastic setting, the problem corresponds to errors-in-variables identification. A modification of the generic problem considered is presented that is a deterministic equivalent to the classical ARMAX identification. The modification is also a mosaic Hankel structured low-rank approximation problem.

Index Terms— **system identification; behavioral approach; missing data; mosaic Hankel matrix; low-rank approximation.**

I. Introduction

System identification aims at deriving a dynamical model $\widehat{\mathscr{B}}$ (*i.e.*, a mathematical description) of a to-be-modeled physical plant from observed data $\mathscr{D}$. The data is typically obtained by sampling and quantization in time-domain. One or more independent measurement experiments can be performed. Each experiment yields a real-valued vector time series. We refer to the individual observations of the variables of the system as *measurement points*. The model postulates a relation among the observed variables, and possibly, some additional unobserved variables. An example of the latter type of model is the auto-regressive moving average exogenous (ARMAX) model. Prior knowledge and/or assumptions about the plant are incorporated in the identification problem by restricting the model to belong to a set of models $\mathscr{M}$, called the *model class*.

An identification problem is a map from data to model:

$$\underset{\mathscr{D}}{\text{data}} \quad \xrightarrow{\text{identification problem}} \quad \underset{\widehat{\mathscr{B}} \in \mathscr{M}}{\text{model}}$$

defined implicitly as a solution to an optimization problem, *i.e.*, the model $\widehat{\mathscr{B}}$ minimizes (among all feasible models) a specified cost function. Different identification problems correspond to different choices of the model class and the cost function.

This conference paper is an abbreviated and updated version of [1], with new material in Sections V and VI-B.

Department ELEC, Vrije Universiteit Brussel, Pleinlaan 2, Building K, B-1050 Brussels, Belgium; Email: imarkovs@vub.ac.be

The objectives of achieving simultaneously

1) "simple" model, and
2) "good" fit of the data by the model

are contradictory. Independent on the data, an arbitrary good fit can be obtained by increasing the model complexity. Typically the model class is used to impose a hard bound on the model complexity and the cost function is used to measure the model-data misfit (lack of fit). It is possible, however, to minimize the model complexity subject to a hard bound on the misfit or, more generally, consider the bi-objective minimization of misfit and complexity.

In exact identification, see, *e.g.*, [2, Chapter 7], the model complexity is minimized subject to the constraint that the model fits the data exactly (zero misfit). If such a model exists in the model class, it is called the *most powerful unfalsified model* (of $\mathscr{D}$ in $\mathscr{M}$) [3]. Exact identification is not a practical identification tool due to the presence of plant disturbances (unobserved variables), measurement noises, discretization and quantization errors. The concept of the most powerful unfalsified model, however, is theoretically important (it is a generalization of the realization problem in system theory) and appears in approximate and stochastic identification problems [4]. Moreover, methods for computing the most powerful unfalsified model lead directly to the very successful class of the subspace identification methods. An exact identification method for data with missing values is presented in [5].

In this paper, we consider the model class of linear time-invariant systems of bounded complexity, defined in Section II. The approximation criterion, which specify the identification problem considered in the paper has the geometric interpretation of the Euclidean distance between the data and the model. In the stochastic setting, this criterion corresponds to errors-invariables system identification [6], *i.e.*, the identification problem considered defines the maximum-likelihood estimator in the errors-invariables setting. In Section V, a modification of the problem for ARMAX system identification is presented. Section III related the identification problem to the weighted mosaic-Hankel structured low-rank approximation. Solution methods for the latter are outlined in Section IV and compared on simulation examples in Section VI.

II. Problem formulation

Model class: bounded complexity linear time-invariant systems

A discrete-time dynamical system $\mathscr{B}$ is a collection of trajectories — q-variables time-series $w : \mathbb{Z} \to \mathbb{R}^q$. The class of finite dimensional linear time-invariant systems with q variables and at most $\mathtt{m}$ inputs is denoted by $\mathscr{L}^q_{\mathtt{m}}$. A linear time-invariant system $\mathscr{B} \in \mathscr{L}^q_{\mathtt{m}}$ admits a representation by constant coefficients difference equation

$$\mathscr{B} = \mathscr{B}(R) := \{\, w \mid R_0 w + R_1 \sigma w + \cdots + R_\ell \sigma^\ell w = 0 \,\}, \quad \text{(DE)}$$

where σ is the shift operator

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

The minimal natural number ℓ, for which there exists an ℓth order difference equation representation for $\mathscr{B}$ is an important invariant of the system, called the *lag*. The number of inputs and the lag specify the complexity of the system in the sense that the dimension of the restriction of $\mathscr{B}$ to the interval $[1,T]$, where $T \geq \ell$, is bounded by $T\mathtt{m} + \ell(q-\mathtt{m})$. The subset of $\mathscr{L}^q_{\mathtt{m}}$ with lag at most ℓ is denoted by $\mathscr{L}^q_{\mathtt{m},\ell}$.

No a priori separation of the variables into inputs and output is made, however, the variables w can always be partitioned into inputs u (free variables) and outputs y (dependent variables) and the system can be represented in the input/state/output representation

$$\mathscr{B} = \mathscr{B}(A,B,C,D,\Pi) := \{\, w = \Pi(u,y) \mid \text{there is } x, \text{ such that } \sigma x = Ax + Bu,\ y = Cx + Du \,\}, \quad \text{(I/S/O)}$$

where Π is a permutation matrix. If $\Pi = I_q$, it will be skipped, *i.e.*, $\mathscr{B}(A,B,C,D) = \mathscr{B}(A,B,C,D,I)$.

The number of inputs $\mathtt{m}$, the number of outputs $\mathtt{p} = q - \mathtt{m}$, and the minimal state dimension $\mathtt{n}$ of an input/state/output representation of $\mathscr{B}$ are invariant of the representation and in particular of the input/output partitioning. The order $\mathtt{n}$ of a state-space representation of a linear time-invariant system $\mathscr{B} = \mathscr{B}(R)$ with lag ℓ and $\mathtt{p}$ outputs is $\mathtt{n} \leq \ell\mathtt{p}$. In the case when the block $P_\ell \in \mathbb{R}^{\mathtt{p}\times\mathtt{p}}$ of $R_\ell = \begin{bmatrix} Q_\ell & -P_\ell \end{bmatrix}$ is nonsingular, $\mathtt{n} = \ell\mathtt{p}$ and $w = (u,y)$ is a possible input/output partition, *i.e.*, Π can be chosen equal to I. This simplifying assumption is made in the rest of the paper. The class of systems with q variables and inputs, order, and lag bounded by, respectively $\mathtt{m}$, $\mathtt{n}$, and ℓ is denoted by $\mathscr{L}^{q,\mathtt{n}}_{\mathtt{m},\ell}$.

Approximation criterion: data-model misfit

We use the behavioral language [7]. The data used for identification is a set

$$w_{\mathrm{d}} = \{\, w^1_{\mathrm{d}}, \ldots, w^N_{\mathrm{d}} \,\} \qquad (w_{\mathrm{d}})$$

of finite trajectories

$$w^k_{\mathrm{d}} = \big(w^k_{\mathrm{d}}(1), \ldots, w^k_{\mathrm{d}}(T_k)\big), \quad \text{where} \quad w^k_{\mathrm{d}}(t) \in \mathbb{R}^q. \qquad (w^k_{\mathrm{d}})$$

The *misfit* (lack of fit) between the data w_{d} and a model $\mathscr{B}$ is measured by the orthogonal distance from w_{d} to $\mathscr{B}$

$$M(w_{\mathrm{d}}, \mathscr{B}) := \min_{\widehat{w}^1, \ldots, \widehat{w}^k \in \mathscr{B}} \sqrt{\textstyle\sum_{k=1}^N \|w^k_{\mathrm{d}} - \widehat{w}^k\|_2^2}. \qquad (M)$$

Missing elements in w_{d} are marked by the symbol `NaN` (not a number). Such elements are excluded from the approximation criterion, *i.e.*, in (M) by definition

$$\mathtt{NaN} - \widehat{w}^k_i(t) = 0, \qquad \text{for all} \quad \widehat{w}^k_i(t) \in \mathbb{R}.$$

The opposite extreme of a missing element is "exact element", *i.e.*, $w^k_{\mathrm{d},i}(t)$ for which the constraint $\widehat{w}^k_{\mathrm{d},i}(t) = w^k_{\mathrm{d},i}(t)$ is imposed in the misfit computation (M).

Identification problem: misfit minimization

The optimal approximate modeling problem considered aims to find a system $\widehat{\mathscr{B}}$ in the model class $\mathscr{L}^q_{\mathtt{m},\ell}$ that best fits the data according to the misfit criterion.

> Given a set of time series w_{d}, specification of exact data, and a complexity specification $(\mathtt{m}, \ell)$, find a system
>
> $$\widehat{\mathscr{B}} := \arg \min_{\mathscr{B} \in \mathscr{L}^q_{\mathtt{m},\ell}} M(w_{\mathrm{d}}, \mathscr{B}). \qquad \text{(SYSID)}$$

Special cases of (SYSID) are static data modeling ($\ell = 0$) and output-only or autonomous system identification ($\mathtt{m} = 0$). The solution approach, described next, leads to an algorithm that covers these special cases. In addition,

1) elements of the given time series w_{d} can be specified as "missing" by passing the symbol `NaN` for their value;
2) elements of the given time series w_{d} can be specified as "exact", in which case they appear unmodified in the approximation $\widehat{w}$; For example, in output error identification problems the variables are a priori partitioned into inputs and outputs, where the input variables are exact while the output variables are perturbed by measurement noise.
3) the approximation $\widehat{w}$ can be constrained to be a trajectory of the model $\mathscr{B}$, generated under a priori fixed initial conditions w_{ini}, see [8], *i.e.*,

   $$\begin{bmatrix} w_{\mathrm{ini}} \\ \widehat{w} \end{bmatrix} \in \mathscr{B}.$$

   (Note that problem (SYSID) identifies the model without prior knowledge about the initial conditions, under which the data w_{d} is generated, *i.e.*, w_{ini} is a free variable.) In identification from impulse or step response data, however, the initial conditions are know exactly. When available this prior information should be taken into account by constraining w_{ini}.

III. Structured low-rank approximation

A. Mosaic-Hankel matrices

It is well known [9] that the realization problem (a special exact identification problem when the data is the impulse response) is closely connected to the rank revealing factorization of a block-Hankel matrix

$$\mathscr{H}_{\ell+1}(h) := \begin{bmatrix} h(1) & h(2) & \cdots & h(T-\ell) \\ h(2) & h(3) & \cdots & h(T-\ell+1) \\ \vdots & \vdots & & \vdots \\ h(\ell+1) & h(\ell+2) & \cdots & h(T) \end{bmatrix},$$

constructed from the first T samples $h = \big(h(1),\dots,h(T)\big)$ of the system's impulse response. More generally (see [1]), if the time-series $w_{\mathrm{d}}^1,\dots,w_{\mathrm{d}}^N$ are exact trajectories of a model $\mathscr{B} \in \mathscr{L}_{\mathtt{m},\ell}$, the mosaic-Hankel matrix (a $1\times N$ block matrix with block-Hankel blocks [10])

$$\mathscr{H}_{\ell+1}(w_{\mathrm{d}}) := \begin{bmatrix}\mathscr{H}_{\ell+1}(w_{\mathrm{d}}^1) & \cdots & \mathscr{H}_{\ell+1}(w_{\mathrm{d}}^N)\end{bmatrix},$$

has rank at most $(\ell+1)\mathtt{m}+\mathtt{n}$.

As in the realization problem, parameters of a state-space representation of the most powerful unfalsified model of w_{d} in $\mathscr{L}_{\mathtt{m},\ell}$ can be obtained from the rank revealing factorization of $\mathscr{H}_{\ell+1}(w_{\mathrm{d}})$ or equivalently from its left kernel:

$$\begin{aligned}\big(w_{\mathrm{d}}^k(1),\dots,w_{\mathrm{d}}^k(T_k-\ell)\big) \in \mathscr{B} \in \mathscr{L}_{\mathtt{m},\ell}^{q,\mathtt{n}}, \quad \text{for } k = 1,\dots,N \\ \iff \quad \operatorname{rank}\big(\mathscr{H}_{\ell+1}(w_{\mathrm{d}})\big) \leq (\ell+1)\mathtt{m}+\mathtt{n}. \quad (*)\end{aligned}$$

B. Dealing with exact and missing data

Let "vec" be a time series vectorization operation, which acting on a trajectory w^k produces a vector of the sequential samples of w^k,

$$p^k = \operatorname{vec}(w^k) := \begin{bmatrix} w^k(1) \\ \vdots \\ w^k(T_k)\end{bmatrix} \in \mathbb{R}^{qT_k}.$$

The identification data w_{d} is represented by the vector

$$p = \operatorname{vec}(w_{\mathrm{d}}) := \begin{bmatrix} \operatorname{vec}(w_{\mathrm{d}}^1) \\ \vdots \\ \operatorname{vec}(w_{\mathrm{d}}^N)\end{bmatrix} \in \mathbb{R}^{q(T_1+\cdots+T_N)}.$$

The inverse operation "vec^{-1}" produces the set of time series w_{d} back from p (and q, T_1, …, T_N):

$$\operatorname{vec}^{-1} : p \mapsto w_{\mathrm{d}}.$$

Define also the element-wise weighted 2-norm

$$\|p\|_v := \sqrt{\textstyle\sum_{i=1}^{n_p} v_i p_i^2},$$

specified by the nonnegative vector $v \in (\mathbb{R}_+ \cup +\infty)^{n_p}$. Finally, let $\mathscr{I}_{\mathrm{exct}}$ be the set of indeces of the exact values of p.

Using the above notation, the approximation criterion (M) can be written as a weighted 2-norm approximation

$$M(w_{\mathrm{d}},\mathscr{B}) := \min_{\widehat{p}} \|p-\widehat{p}\|_v \quad \text{subject to} \quad \operatorname{vec}^{-1}(\widehat{p}) \subset \mathscr{B},$$

with zero weights assigned to the missing data and infinite weights assigned to the exact data, *i.e.*,

$$v_i = \begin{cases} 0 & \text{if } p_i \text{ is } \texttt{NaN} \text{ (missing data)},\\ \infty & \text{if } i \in \mathscr{I}_{\mathrm{exct}} \text{ (exact data)},\\ 1 & \text{otherwise (noisy data)}.\end{cases}$$

Indeed, for the misfit to be finite, the equality constraints $\widehat{p}_i = p_i$ must hold for all $i \in \mathscr{I}_{\mathrm{exct}}$.

C. Structured low-rank approximation

By $(*)$, the identification problem (SYSID) is equivalent to the element-wise weighted mosaic-Hankel structured low-rank approximation problem

$$\begin{aligned}&\text{minimize} \quad \text{over } \widehat{p} \quad \|p-\widehat{p}\|_v^2 \\ &\text{subject to} \quad \operatorname{rank}\big(\mathscr{H}_{\ell+1}\big(\operatorname{vec}^{-1}(\widehat{p})\big)\big) \leq r,\end{aligned} \tag{SLRA}$$

where $r = (\ell+1)q - \mathrm{p}$.

IV. Solution method

Problem (SLRA) is in general a nonconvex optimization problem. In this section, we review a local optimization method based on a kernel representation of the rank constraint

$$\begin{aligned}\operatorname{rank}\big(\mathscr{H}_{\ell+1}(\widehat{w})\big) \leq r \quad \iff \quad R\mathscr{H}_{\ell+1}(\widehat{w}) = 0 \\ \text{and} \quad R \in \mathbb{R}^{\mathrm{p}\times(\ell+1)q} \text{ is full row rank (f.r.r.).}\end{aligned} \tag{KER}$$

The matrix R in the right-hand-side of (KER) is related to the parameters $R_0, R_1, \dots, R_\ell$ of the difference equation representation (DE) of the exact model for $\widehat{w}$ as follows:

$$R = \begin{bmatrix} R_0 & R_1 & \cdots & R_\ell\end{bmatrix}, \qquad \text{where} \quad R_i \in \mathbb{R}^{\mathrm{p}\times q}.$$

Using the kernel representation (KER), the structured low-rank approximation problem (SLRA) is rewritten in the following equivalent form

$$\begin{aligned}&\text{minimize} \quad \text{over } \widehat{p} \text{ and } R \in \mathbb{R}^{\mathrm{p}\times(\ell+1)q} \quad \|p-\widehat{p}\|_v^2 \\ &\text{subject to} \quad R\mathscr{H}_{\ell+1}(\widehat{p}) = 0 \quad \text{and} \quad R \text{ is f.r.r.}\end{aligned} \tag{SLRA$_R$}$$

Problem (SLRA$_R$) is a nonlinear least squares problem, which, separable in the optimization variables $\widehat{p}$ and R. In particular, the variable $\widehat{p}$ can be eliminated by analytically minimizing over it. This reduces (SLRA$_R$) to the equivalent problem:

$$\text{minimize} \quad \text{over f.r.r. } R \in \mathbb{R}^{\mathrm{p}\times(\ell+1)q} \quad M(R), \tag{OUTER}$$

where

$$M(R) := \min_{\widehat{p}} \|p-\widehat{p}\|_v^2 \quad \text{s.t.} \quad R\mathscr{S}(\widehat{p}) = 0. \tag{INNER}$$

The computation of $M(R)$ for given R is refered to as the *inner minimization* problem and the minimization (OUTER) of the function M over R is referred to as *outer minimization* problem.

The inner minimization problem (INNER) is a linear least norm problem and admits an analytic solution. In [11], it is sown how this problem can be solved in the presence of exact ($v_i = +\infty$) and missing ($v_i = 0$) values. In [12], it is shown that fast $O(T)$ evaluation of M and its derivatives can be performed for mosaic-Hankel-like structured matrices.

The approach for solving (SLRA$_R$), based on elimination of $\widehat{p}$, is closely related to the variable projection method [13]. In [13], however, an explicit function $\widehat{b} = A(\theta)x$, where x is unconstrained, is considered, while in the context of the structured low-rank approximation problem, an implicit function (relation) $R\mathscr{S}(\widehat{p}) = 0$ is considered, where the variable R is constrained to have full row rank. This fact

requires new type of algorithms, where the nonlinear least squares problem is an optimization problem on a Grassmann manifold, see [14], [15].

In (OUTER), the cost function M is minimized over the set of full row rank matrices R. Note that, M depends only on the space spanned by the rows of R. In order to find a minimum of M, the search space in (OUTER) can be replaced by the matrices satisfying the constraint

$$RR^\top = I_{\mathrm{p}},$$

or

$$R = \begin{bmatrix} X & I_{\mathrm{p}} \end{bmatrix} \Pi,$$

where X is a free variable and Π is a $q \times q$ permutation matrix. (In a system theoretic setting, Π defines an input/output partitioning of the variables.)

A software package for mosaic-Hankel structured low-rank approximation is presented in [16]. The Levenberg-Marquardt algorithm [17] implemented in the GNU Scientific Library [18], is used for the solution of the nonlinear least squares problem. This package is used in [1] for system identification.

V. Latency minimization: ARMAX system identification

The classical setting for system identification is the ARMAX one [19], [20]. An ARMAX model is a linear time-invariant system driven by an unobserved disturbance e as well as the observed input u:

$$\mathscr{B}_{\mathrm{ext}}(P,Q,E) = \left\{ \begin{bmatrix} e \\ u \\ y \end{bmatrix} \;\middle|\; P(\sigma)y = Q(\sigma)u + E(\sigma)e \right\}.$$

Here P, Q, and E are polynomial matrices parameterizing the model. The disturbance e is modeled as a zero mean, white noise process. The aim of the ARMAX identification problem is to find the dynamical relation between the disturbance and the output (the "noise dynamics", defined by E and P) as well as the dynamical relation between the inputs and the output (the "signal dynamics", defined by Q and P).

The identification problem (SYSID) and the solution approach based on the structured low-rank approximation problem (SLRA) do not included the ARMAX setting as a special case. Vice verse, the ARMAX setting is considered unrelated to the identification problem (SYSID), the stochastic equivalent of which is errors-in-variables identification.

In [21], [22] a deterministic equivalent of the ARMAX problem is proposed, in which the unobserved input e is treated as a deterministic latent variable rather than stochastic process. Using the notation in this paper, the deterministic ARMAX identification problem is

$$\begin{aligned} &\text{minimize} && \text{over } \widehat{\mathscr{B}}_{\mathrm{ext}} \text{ and } \widehat{e} \quad \|\widehat{e}\|_2 \\ &\text{subject to} && \begin{bmatrix} \widehat{e} \\ w_{\mathrm{d}} \end{bmatrix} \in \widehat{\mathscr{B}}_{\mathrm{ext}} \in \mathscr{L}_{q,\ell} \cap \mathscr{C}, \end{aligned} \tag{ARMAX}$$

where

$$\mathscr{C} = \{ \mathscr{B}(P,Q,E) \mid E_{\ell+1} = I_{\mathrm{p}} \text{ and } P_{\ell+1} = I_{\mathrm{p}} \}. \tag{1}$$

Given, data w_{d} and model $\widehat{\mathscr{B}}_{\mathrm{ext}} \in \mathscr{L}_{q,\ell}$, the quantity

$$L(w_{\mathrm{d}}, \mathscr{B}_{\mathrm{ext}}) := \min_{\widehat{e}} \quad \|\widehat{e}\|_2 \quad \text{s.t.} \quad \begin{bmatrix} \widehat{e} \\ w_{\mathrm{d}} \end{bmatrix} \in \widehat{\mathscr{B}}_{\mathrm{ext}} \tag{L}$$

is called the latency of w_{d} with respect of $\mathscr{B}_{\mathrm{ext}}$. The latency computation in the deterministic setting corresponds to the prediction errors or likelihood computation in the stochastic setting. In both cases Kalman filtering or smoothing algorithms are used. A deterministic equivalent of the Kalman filter is considered in [23], [24], [25]

VI. Simulation examples

A. Errors-in-variables setup

In this section, we show examples of identification problems with missing data. The data w_{d} is generated in the errors-in-variables setting

$$w_{\mathrm{d}} = \bar{w} + \widetilde{w},$$

where the true data $\bar{w}$ is a trajectory of a true model $\bar{\mathscr{B}} \in \mathscr{L}^q_{\mathrm{m},\ell}$ and the measurement noise $\widetilde{w}$ is white zero mean normally distributed with covariance matrix that is a multiple of the identity. The true model is single-input single-output linear time-invariant system $\bar{\mathscr{B}} = \mathscr{B}(\bar{R})$, where $\bar{R} = \begin{bmatrix} -\bar{Q} & \bar{P} \end{bmatrix}$ with

$$\begin{aligned} \bar{P}(z) &= z^2 - 1.4z + 0.7, \\ \bar{Q}(z) &= -z - 0.3. \end{aligned} \tag{PAR}$$

Samples of w_{d} are missing at moments of time $t \in \mathscr{T}_{\mathrm{m}}$, where $\mathscr{T}_{\mathrm{m}}$ is specified in the examples.

An alternative method for solving the errors-in-variables identification problem with missing data is proposed in [26]. This method uses a frequency domain approach [27]. A Matlab implementation of the algorithm (called below `sysid`) was kindly provided by the authors and is used below for verification of the results obtained with the function `ident`, based on the structured low-rank approximation method of Section IV.

The identified models $\widehat{\mathscr{B}}$ are evaluated by the angle

$$\angle(\bar{P}, \widehat{P}) = \cos^{-1}\left(\frac{\bar{P}^\top \widehat{P}}{\|\bar{P}\| \|\widehat{P}\|} \right), \tag{ERR}$$

between the true $\bar{P}$ and estimated $\widehat{P}$ model parameter vectors.

The simulation parameters in the experiments are the number of samples T, the set of missing values $\mathscr{T}_{\mathrm{m}}$, and the noise standard deviation. The reported results show the approximation error for the compared methods and for different noise levels. Three experiments are done for different distribution of the missing values: sequential, periodic, and random. Both input and output values are missing. A `NaN` value in the table of results indicates that the corresponding method fails in this case.

In an example with $T = 100$ and sequential missing data in the interval $[30, 70]$, the `ident` and `sysid` functions achieve comparable accuracy (see Table I). The `ident` function is 9 time faster than `sysid`. Similar results (see Table II) are obtained for small noise levels in the case of periodic missing data in the interval $[30, 70]$ with period 3.

TABLE I

ESTIMATION ERRORS (ERR) IN THE ERRORS-INVARIABLES SETUP, WITH SEQUENTIAL MISSING INPUT AND OUTPUT SAMPLES IN THE INTERVAL $[30,70]$, $T = 100$.

noise level	0	0.0025	0.0050	0.0075	0.0100
`ident`	10^{-8}	0.0012	0.0078	0.0017	0.0067
`sysid`	0	0.0012	0.0078	0.0014	0.0065

TABLE II

ESTIMATION ERRORS (ERR) IN THE ERRORS-INVARIABLES SETUP, WITH PERIODIC MISSING INPUT AND OUTPUT SAMPLES IN THE INTERVAL $[30,70]$ WITH PERIOD 3, $T = 100$.

noise level	0	0.0025	0.0050	0.0075	0.0100
`ident`	10^{-8}	0.0014	0.0058	0.0077	0.0068
`sysid`	0	0.0014	0.0073	0.0082	0.0077

In this example, the `ident` function is 5 time faster than `sysid`. Finally, results for a simulation example with $T = 1000$ data points from which 600 are randomly missing are shown in Table III.

TABLE III

ESTIMATION ERRORS (ERR) IN IN THE ERRORS-INVARIABLES SETUP, WITH RANDOMLY DISTRIBUTED MISSING INPUT AND OUTPUT SAMPLES, $T = 1000$, 600 MISSING.

noise level	0	0.0025	0.0050	0.0075	0.0100
`ident`	10^{-5}	0.0029	0.0087	0.0028	0.0123
`sysid`	`NaN`	`NaN`	`NaN`	`NaN`	`NaN`

B. ARMAX identification

In this section, we validate on a numerical examples the results of the method described in Section IV, applied for solving the deterministic formulation (ARMAX) of the ARMAX identification problem. Also we show results for ARMAX identification with missing data, corresponding to the ones in Section VI-A in the errors-in-variables setup.

The standard approach for ARMAX system identification is prediction error minimization. Similarly to the method for structured low-rank approximation, presented in Section IV, the prediction error minimization methods are local optimization methods and have a double minimization structure. The inner minimization is the prediction error (or likelihood) evaluation, which corresponds to the misfit computation in the structured low-rank approximation problem. The outer minimization is a nonlinear least squares problem over the model parameters and the initial conditions.

Note 1 (Initial conditions). There is an important difference in the way the initial conditions are taken into account in the prediction error methods and the method of Section IV. In the prediction error methods, the initial conditions are optimization variables in the outer level—the nonlinear least squares minimization, *i.e.*, they are treated as additional model parameters. In the variable projection method of Section IV, the initial conditions are eliminated in the inner minimization problem—the linear least norm problem. Thus the minimization in the outer level is over a fewer optimization variables. This may lead to improved efficiency and robustness of the method described in this paper.

We use the functions `pe` and `armax` from the System Identification Toolbox [28] of Matlab for, respectively, evaluation of the prediction errors and system identification by prediction error minimization. The function `armax` is not directly applicable for identification with missing data and is used in combination with the function `misdata` from the System Identification Toolbox [28] of Matlab. Contrary to the case of no missing data, the combination of `misdata` and `armax` is not an optimal estimation method. As explained in Section IV, the structured low-rank approximation algorithm deals with the missing values by minimizing a weighted cost function with zero weights assigned to the missing values. Therefore, no preprocessing step or adaptation is needed and the method is optimal.

The data w_{d} used in the simulation examples is generated by a "true" ARMAX system $\mathscr{B}_{\mathrm{ext}}(\bar{P},\bar{Q},\bar{E})$, with parameters $\bar{Q}$ and $\bar{P}$ given in (PAR) and $\bar{E}(z) = z + 0.5$. The number of samples is $T = 100$. The input u is random and the system starts from zero initial conditions. The disturbance e is a zero mean white normally distributed signal. Table IV shows the estimation errors in the case of no missing values. The latency $L(w_{\mathrm{d}},\widehat{\mathscr{B}}_{\mathrm{ext}})$ of both models (computed with both the SLRA package [16] and the function `pe`) is the same. This result is a numerical confirmation that the deterministic latency minimization corresponds to the prediction error minimization. As shown in Tables V and VI, the results obtained with the combination of the `misdata` and `armax` methods are suboptimal.

TABLE IV

ESTIMATION ERRORS (ERR) IN THE ARMAX SETUP WITHOUT MISSING VALUES.

noise level	0	0.025	0.05	0.075	0.1
`armax`	0	0.0041	0.0056	0.0050	0.0066
`slra`	0	0.0041	0.0056	0.0051	0.0063

TABLE V

ESTIMATION ERRORS (ERR) IN THE ARMAX SETUP WITH MISSING INPUT AND OUTPUT SAMPLES IN THE INTERVAL $[30,70]$, $T = 100$.

noise level	0	0.025	0.05	0.075	0.1
`misdata+armax`	0	0.0125	0.0113	0.0083	0.0457
`slra`	0.0216	0.0248	0.0137	0.0659	0.6532

TABLE VI

ESTIMATION ERRORS (ERR) IN THE ARMAX SETUP WITH PERIODIC MISSING INPUT AND OUTPUT SAMPLES IN THE INTERVAL $[30,70]$, WITH PERIOD 3, $T = 100$.

noise level	0	0.025	0.05	0.075	0.1
`misdata+armax`	0	0.2205	0.0101	0.0143	0.1584
`slra`	0.0143	0.9133	0.9196	0.9604	0.9891

VII. Conclusions

The paper presented a generic problem for system identification in the behavioral setting. The data consists of multiple time-series. Exact and missing values can be present in arbitrary combination of variables (inputs and outputs) and time instances. The identification problem is equivalent and is solved as a mosaic Hankel structured low-rank approximation with an element-wise weighted 2-norm cost function. Zero weights are assigned to the missing values and infinite weights are assigned to the exact values of the given data. A method based on local optimization is presented and tested on simulation examples. The computational cost of the algorithm is linear in the number of data points. Despite of its generality and flexibility, the developed software is functionally equivalent to and computationally faster than the existing alternatives. A modification of the problem and solution method for ARMAX system identification is also presented and illustrated on examples.

Acknowledgements

The structured low-rank approximation package [16], used in the implementation of the identification method, is a joint work with Konstantin Usevich. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement number 258581 "Structured low-rank approximation: Theory, algorithms, and applications".

References

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

[2] I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor, *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*, ser. Monographs on Mathematical Modeling and Computation. SIAM, March 2006, no. 11.

[3] J. C. Willems, "From time series to linear system—Part II. Exact modelling," *Automatica*, vol. 22, no. 6, pp. 675–694, 1986.

[4] I. Markovsky, *Low Rank Approximation: Algorithms, Implementation, Applications*. Springer, 2012.

[5] ——, "Exact identification with missing data," in *Proc. of the 52nd IEEE Conference on Decision and Control*. Firenze, Italy, 2013.

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

[7] J. C. Willems, "From time series to linear system—Part I. Finite dimensional linear time invariant systems, Part II. Exact modelling, Part III. Approximate modelling," *Automatica*, vol. 22, 23, pp. 561–580, 675–694, 87–115, 1986, 1987.

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

[9] E. D. Sontag, *Mathematical control theory: Deterministic finite dimensional systems*. Springer-Verlag, 1990.

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

[11] I. Markovsky and K. Usevich, "Structured low-rank approximation with missing data," *SIAM J. Matrix Anal. Appl.*, pp. 814–830, 2013.

[12] K. Usevich and I. Markovsky, "Variable projection for affinely structured low-rank approximation in weighted 2-norms," *J. Comput. Appl. Math.*, 2013.

[13] 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.

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

[15] P.-A. Absil, R. Mahony, R. Sepulchre, and P. V. Dooren, "A Grassmann–Rayleigh quotient iteration for computing invariant subspaces," *SIAM Review*, vol. 44, no. 1, pp. 57–73, 2002.

[16] I. Markovsky and K. Usevich, "Software for weighted structured low-rank approximation," *J. Comput. Appl. Math.*, 2013.

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

[18] M. Galassi *et al.*, "GNU scientific library reference manual," http://www.gnu.org/software/gsl/.

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

[20] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1999.

[21] P. Lemmerling and B. De Moor, "Misfit versus latency," *Automatica*, vol. 37, pp. 2057–2067, 2001.

[22] J.C. Willems, *Thoughts on system identification*. In Control of Uncertain Systems: Modelling, Approximation, and Design, Springer, vol. 329, pp. 389–416, 2006.

[23] J. C. Willems, "Deterministic least squares filtering," *J. Econometrics*, vol. 118, pp. 341–373, 2004.

[24] I. Markovsky, J. C. Willems, and B. De Moor, "Continuous-time errors-in-variables filtering," in *Proc. 41st Conf. on Decision and Control*, Las Vegas, NV, 2002, pp. 2576–2581.

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

[26] R. Pintelon and J. Schoukens, "Frequency domain system identification with missing data," *IEEE Trans. Automat. Control*, vol. 45, no. 2, pp. 364–369, Feb. 2000.

[27] ——, *System Identification: A Frequency Domain Approach*, 2nd ed. Piscataway, NJ: IEEE Press, 2012.

[28] L. Ljung, *System identification toolbox: User's guide*, The MathWorks, 2013.