

On the most powerful unfalsified model for data with missing values (special issue JCW)

Ivan Markovsky
Department ELEC
Vrije Universiteit Brussel (VUB)
Pleinlaan 2, Building K
B-1050 Brussels, Belgium
Fax: +32-2-629.28.50
ivan.markovsky@vub.ac.be

Abstract

The notion of the most powerful unfalsified model plays a key role in system identification. Since its introduction in the mid 80's, many methods have been developed for its numerical computation. All currently existing methods, however, assume that the given data is a *complete* trajectory of the system. Motivated by the practical issues of data corruption due to failing sensors, transmission lines, or storage devices, we study the problem of computing the most powerful unfalsified model from data with missing values. We do not make assumptions about the nature or pattern of the missing values apart from the basic one that they are a part of a trajectory of a linear time-invariant system. The identification problem with missing data is equivalent to a Hankel structured low-rank matrix completion problem. The method proposed constructs rank deficient complete submatrices of the incomplete Hankel matrix. Under specified conditions the kernels of the submatrices form a nonminimal kernel representation of the data generating system. The final step of the algorithm is reduction of the nonminimal kernel representation to a minimal one. Apart from its practical relevance in identification, missing data is a useful concept in systems and control. Classic problems, such as simulation, filtering, and tracking control can be viewed as missing data estimation problems for a given system. The corresponding identification problems with missing data are "data-driven" equivalents of the classical simulation, filtering, and tracking control problems.

Keywords: most powerful unfalsified model, exact system identification, subspace methods, missing data, low-rank matrix completion.

1 Introduction

1.1 Context and aim of the paper

The behavioral approach to systems and control was developed from "a need to put a clear and rational foundation under the problem of obtaining models from time series" [27, page 561]. One of the key ideas that came out from the original work [27, 28, 29] of Jan Willems is the notion of the "most powerful unfalsified model", or MPUM for short. The MPUM is "unfalsified" in the sense that it is an exact model for the given data and "most powerful" in the sense that it is the least complicated exact model. Thus, the MPUM is an optimal exact model for the data.

A candidate model $\hat{\mathcal{B}}$ for the data w_d is unfalsified if w_d is a trajectory of $\hat{\mathcal{B}}$. In the behavioral setting this fact is conveniently written as $w_d \in \hat{\mathcal{B}}$. (By definition the model is the set of all valid trajectories.) Restricting to the class of linear time-invariant models and assuming that the number of the input variables is a priori fixed, the complexity of the model can be quantified by its order or by its lag. Let \mathcal{L}_m be the set of linear time-invariant systems with at most m inputs. The MPUM for the data w_d in the model class \mathcal{L}_m is defined as

$$\mathcal{B}_{\text{mpum}}(w_d) := \arg \underbrace{\min_{\hat{\mathcal{B}} \in \mathcal{L}_m} \text{lag}(\hat{\mathcal{B}})}_{\text{most powerful}} \quad \text{subject to} \quad \underbrace{w_d \in \hat{\mathcal{B}}}_{\text{unfalsified model}} . \quad (\text{MPUM})$$

Apart from defining the notion of the MPUM, in [28], Jan Willems developed algorithms that implement the mapping $w_d \mapsto \mathcal{B}_{\text{mpum}}(w_d)$. These algorithms motivated the development of the subspace identification methods. The so-called "deterministic subspace identification" problem, see [24, Chapter 2] and [18, Chapter 7], is the problem of computing a state space representation of an exact model from data. Unlike the methods of [28], the subspace identification methods assume a priori given upper bound of the lag or the order of the model. If this bound is over specified, a nonminimal representation of $\mathcal{B}_{\text{mpum}}(w_d)$ is computed. Subsequently, it is reduced to a minimal one. Thus instead of optimizing over the model complexity, the subspace methods use model reduction in order to find the MPUM.

The class of the subspace methods was generalized to approximate identification in the ARMAX [24, 25] and errors-in-variables [3] settings, identification in closed-loop [2], identification of dissipative and lossless systems [6, 20], and other constrained identification problems. The subspace methods are practically successful and are still developed theoretically, generalized to new problems, and improved computationally.

In this paper, we consider the exact (deterministic) identification problem in the case of data with missing values. Apart from the preliminary results [13] by the author, currently there are no subspace methods that address this problem. We do not make assumptions about the nature or pattern of the missing values apart from the basic one that they are a part of a valid trajectory of a linear time-invariant system with a given number of inputs and bounded lag. The missing elements can be both inputs and outputs of the system under some given input/output partitioning of the variables and they can appear in any pattern in time: periodically, randomly, or in blocks of consecutive time samples.

1.2 Literature review and contribution

Most of the currently existing literature on identification with missing data addresses special cases, such as specific patterns of occurrence of the missing data, or uses heuristics for estimation of the missing data, such as interpolation methods, followed by classical identification from the completed data. Three important special cases and three state-of-the-art methods that solve the general problem are reviewed next.

Special cases

The following special identification problems with missing data were considered in the literature:

- partial realization problem,
- missing input and output values in a single block, and
- missing values in the output only.

The partial realization problem is an exact identification problem from data consisting of the first few samples of the impulse response. This problem can be posed and solved as an extension of the given samples of the impulse response, *i.e.*, estimation of the missing output values, after the given ones. Kalman derived an analytical solution [7, 8] for this problem. This solution, however, does not generalize to other patterns of missing data.

Another special identification problem with missing data considered in the literature [21] is the problem when missing are $w_d(t), w_d(t+1), \dots, w_d(t + \text{lag}(\mathcal{B}))$. In this case, the identification problem with missing data is equivalent to identification from two independent data sets: $w_d^1 = (w_d(1), \dots, w_d(t-1))$ and $w_d^2 = (w_d(t + \text{lag}(\mathcal{B}) + 1), \dots, w_d(T))$, where T is the number of samples of w_d . This result also does not generalize to other patterns of missing values.

The special case when the missing data is restricted to the output variables only can be handled by the classical prediction error identification methods [22, 10]. The predictor is used to estimate the missing output values from the inputs, the current guess of the model and the initial conditions.

Optimization-based methods

The general identification problem with missing data can be approached by choosing a representation of the model and optimizing the complexity over the model parameters and the missing values, subject to the constraint that the completed data is a trajectory of the system. This leads to a nonconvex optimization problem. Three classes of methods that use this approach are:

- modification of the classical prediction error methods,
- methods developed in the structure low-rank approximation setting [16, 14, 15], and
- convex relaxation methods based on the nuclear norm heuristic.

All these methods are designed for estimation from noisy as well as missing data.

The approach using the prediction error methods for missing data estimation in the outputs was recently generalized in [26] to missing values of both inputs and outputs. Standard nonlinear local optimization methods are used. These methods require initial values for the optimization variables (model parameters and missing values) and the results depend on their closeness to a "good" locally optimal solution. Similar in spirit but different in implementation details [23, 17, 12] are the methods developed in the structure low-rank approximation setting.

An approach that gained popularity lately due to its success in compressive sensing is relaxation of the problem to a convex one by using the nuclear norm in lieu of the rank [4]. In [9], system identification with missing data is handled by 1) completion of the missing data using the nuclear norm heuristic (this step requires solution of a convex optimization problem), and 2) identification of a model parameter from the completed sequence using classical subspace identification methods. In the context of identification from noisy data, the optimization problem on step 1 involves a trade-off between the model complexity and the model accuracy. This trade-off is set by a user defined hyper-parameter. In the context of the exact identification problem considered in this paper, there is no trade-off parameter, see Section 5.1.

Contribution and organization of the paper

Our main contribution is a subspace type method for exact identification of a linear time-invariant system from data with missing values. Compared with the method based on the nuclear norm heuristic, the subspace method uses only linear algebra operations such as kernel computation and solution of linear systems of equations, which makes it computationally more efficient.

Another contribution of the paper is using the missing data estimation methods for solving systems and control related problems, besides model identification. More specifically, we solve simulation and output tracking control problems by Algorithms 1 and 2, presented in Section 4. These examples show that missing data estimation is a unifying tool for systems and control related problems.

The paper is organized as follows. In Section 2 we define the notation being used. Section 3 defines formally the problems considered:

1. verification when a sequence w_d with missing data is a trajectory of a *given* linear time-invariant system \mathcal{B} , and
2. methods for computing $\mathcal{B}_{\text{mpum}}(w_d)$ from a sequence w_d with missing values.

The solution to these problems are presented in Sections 4 and 5, respectively. The methods proposed are illustrated on examples and their advantages and disadvantages are compared. Conclusions and directions for future work are given in Section 6.

2 Preliminaries

Missing data values are denoted by the symbol NaN (“not a number”). The extended set of real numbers \mathbb{R}_e is the union of the set of the real numbers \mathbb{R} and the symbol NaN:

$$\mathbb{R}_e := \mathbb{R} \cup \text{NaN}.$$

The data for the considered problem is a q -variate sequence

$$w_d = (w_d(1), \dots, w_d(T)), \quad \text{where } w_d(t) \in \mathbb{R}_e^q.$$

The sequence w_d is parameterized by the $Tq \times 1$ vector

$$\text{vec}(w_d) := \text{col}(w_d(1), \dots, w_d(T)).$$

The subvector of a vector a with indexes in \mathcal{I} is denoted by $a_{\mathcal{I}}$. Similarly, for a matrix A , $A_{\mathcal{I}}$ is the submatrix formed by the rows with indexes in the set \mathcal{I} , and $A_{\mathcal{I},\mathcal{J}}$ is the submatrix of A with elements a_{ij} , such that $i \in \mathcal{I}$, $j \in \mathcal{J}$. The set of integers from i to j is denoted by $i:j$. The set of the indexes of the missing elements of the data vector $\text{vec}(w_d)$ is denoted by \mathcal{I}_m and the set of the indexes on the given elements by \mathcal{I}_g .

We are interested in sequences that are trajectories of discrete-time linear time-invariant dynamical systems. A linear time-invariant dynamical system \mathcal{B} with q variables is a set that can be represented as the kernel of a polynomial operator,

$$\mathcal{B} = \ker(R(\sigma)) := \{w \mid R(\sigma)w = 0\}, \quad (\text{KER})$$

where

$$R(z) = R_0 + R_1z + \cdots + R_\ell z^\ell$$

is a polynomial matrix with q columns and σ is the backwards shift operator

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

The minimal natural number ℓ , for which there exists an ℓ th order difference equation representation for \mathcal{B} is an invariant of the system, called the *lag*.

Assuming that the linear time-invariant system \mathcal{B} is controllable, it admits an image representation

$$\mathcal{B} = \text{image}(M(\sigma)) := \{w \mid \text{there is } v, \text{ such that } w = M(\sigma)v\}, \quad (\text{IM})$$

where

$$M(z) = M_0 + M_1z + \cdots + M_\ell z^\ell$$

is a polynomial matrix with q rows.

The variables can always be partitioned element-wise into inputs u and outputs y . The number of inputs m and the number of outputs $p = q - m$ are representation invariant. With some loss of generality, we assume that the first m elements of w are inputs and the remaining elements are outputs, *i.e.*, $w = \begin{bmatrix} u \\ y \end{bmatrix}$. Then, the linear time-invariant system \mathcal{B} admits the classical input/state/output representation

$$\mathcal{B} = \mathcal{B}_{i/s/o}(A, B, C, D) := \left\{ w = \begin{bmatrix} u \\ y \end{bmatrix} \mid \text{there is } x, \text{ such that } \sigma x = Ax + Bu \text{ and } y = Cx + Du \right\}. \quad (\text{I/S/O})$$

The set of finite T -sampled long trajectories of \mathcal{B} is denoted by \mathcal{B}_T . The statement " w_d is a finite T -samples long trajectory of the system \mathcal{B} " is concisely written as $w_d \in \mathcal{B}_T$. If w_d contains missing values, the assertion $w_d \in \mathcal{B}_T$ means that the missing values w_{d,\mathcal{I}_m} can be completed, so that the completed sequence, say \hat{w} , is a finite T -samples long trajectory of \mathcal{B} . The completion $\hat{w}_{\mathcal{I}_m}$ should exist but need not be unique.

3 Problem formulation

In this section, we define formally the problems solved in the paper. Section 3.1 poses the problem of verifying when a given sequence with missing values is a trajectory of a given model. Solving this problem requires completion of the missing values, which is a task of high practical as well as theoretical interest. After posing the missing values completion problem with a given model, we define in Section 3.2 the more difficult problem of completing a given sequence without knowledge of the data generating model, apart from the prior knowledge that it is linear time-invariant and has bounded lag. This is the problem of exact identification from data with missing values.

3.1 Completing of a sequence with a given model

Problem 1 (Missing values estimation with a given model). Given a linear time-invariant system $\mathcal{B} \in \mathcal{L}_m$ and a sequence $w_d \in (\mathbb{R}_c^q)^T$ with with missing values $w_{d,\mathcal{I}_m} = \text{NaN}$ and given values w_{d,\mathcal{I}_g} , check if $w_d \in \mathcal{B}_T$, *i.e.*, find a complete sequence $\hat{w} \in (\mathbb{R}^q)^T$ that agrees with the given data $\hat{w}_{\mathcal{I}_g} = w_{d,\mathcal{I}_g}$ and is a trajectory of the system \mathcal{B} , or assert that no such sequence exists.

Well known special cases of Problem 1 are:

- simulation—given input and initial conditions, find the corresponding output, and
- output matching control—given initial conditions and output, find an input that generates the specified output.

Initial conditions can be specified in a representation free manner by a prefix trajectory, *i.e.*, an initial part of w_d that is at least $\text{lag}(\mathcal{B})$ samples long. Following a tradition in the subspace identification literature, we refer to the prefix trajectory as the "past" and to the remaining part of the trajectory as the "future". In the simulation problem, missing is the future output. In the output matching control, missing is the future input.

In order to distinguish Problem 1 from the identification problem, defined next, we refer to Problem 1 as the "estimation problem".

3.2 Exact identification from trajectory with missing values

Problem 2 (Exact system identification from trajectory with missing values). Given a sequence $w_d \in (\mathbb{R}_e^q)^T$ and a model class \mathcal{L}_m ,

$$\text{minimize over } \hat{\mathcal{B}} \in \mathcal{L}_m \text{ and } \hat{w} \in (\mathbb{R}^q)^T \quad \text{lag}(\hat{\mathcal{B}}) \quad \text{subject to } \hat{w}_{\mathcal{I}_g} = w_{d,\mathcal{I}_g} \quad \text{and} \quad \hat{w} \in \hat{\mathcal{B}}_T. \quad (\text{P})$$

Without missing values, (P) coincides with the definition (MPUM) of $\mathcal{B}_{\text{mpum}}(w_d)$. With missing values, the test $\hat{w} \in \hat{\mathcal{B}}$ requires completion of the missing values (Problem 1). If a completion exists for a given model $\hat{\mathcal{B}}$, then this model is feasible. The aim of Problem 2 is to find the least complicated feasible model. This is a generalization of the notion of the MPUM for data with missing values.

The method developed in the paper is not based on solving the nonconvex optimization problem (P). Instead, it derives a nonminimal kernel representation from the data with missing values and reduces it to a minimal one. The derivation of the nonminimal representation is done by finding completely specified submatrices of the incomplete Hankel matrix of the data and computing their kernels.

We refer to Problem 2 as the "identification problem".

Note 3 (Multiple trajectories). Problems 1 and 2 can be extended straightforwardly to data consisting of multiple trajectories with missing data.

4 Missing values completion with a given model

After defining the estimation and identification problems, we proceed to describe methods for their solution. In this section, we consider the estimation problem—missing values completion with a given model. Two algorithms are proposed: Subsection 4.1 presents an algorithm using an image representation of the model, and Subsection 4.2 presents an algorithm using an input/state/output representation. Both algorithms complete the sequence or prove that the sequence can not be completed by solving or establishing inconsistency of a system of linear equations. Finally, subsection 4.3 shows how the algorithms solve as special cases the simulation and output tracking control problems.

4.1 Using image representation

First, we give a solution of Problem 1 in terms of an image representation of the system. Computationally, the solution is checking solvability of a linear system of equations with a low-triangular banded block-Toeplitz coefficients matrix

$$\mathcal{T}_T(M) := \begin{bmatrix} M_0 & & & & & \\ & M_1 & \ddots & & & \\ & \vdots & \ddots & & & \\ & M_\ell & & M_0 & & \\ & & & & M_1 & \\ & & & & & \vdots \\ & & & & & & M_\ell \end{bmatrix}.$$

Proposition 4. Let \mathcal{B} be a controllable linear time-invariant system with an image representation $\mathcal{B} = \text{image}(M(\sigma))$. The sequence $w_d \in (\mathbb{R}_e^q)^T$ with missing values $w_{d,\mathcal{I}_m} = \text{NaN}$ and given values w_{d,\mathcal{I}_g} is a trajectory of \mathcal{B} if and only if the system of linear equations

$$w_{d,\mathcal{I}_g} = \mathcal{I}_{\mathcal{I}_g}(M)v \quad (v)$$

has a solution \hat{v} , in which case a completion of the missing values is given by

$$\hat{w}_{\mathcal{I}_m} = \mathcal{I}_{\mathcal{I}_m}(M)\hat{v}. \quad (\hat{w})$$

Proof. Since $\mathcal{B}_T = \text{image}(\mathcal{I}_T(M))$, \hat{w} is a trajectory of \mathcal{B} . Moreover, by (v), \hat{w} interpolates the given data w_{d,\mathcal{I}_g} . Therefore, \hat{w} solves Problem 1. \square

The solution method presented in Proposition 4 is summarized in Algorithm 1.

Algorithm 1 Missing values completion with a model specified by an image representation.

Input: Polynomial matrix $M(z)$ defining an image representation $\mathcal{B} = \text{image}(M(\sigma))$ of the system and a sequence $w_d \in (\mathbb{R}_e^q)^T$.

- 1: **if** (v) has solution **then**
 - 2: $w_d \in \mathcal{B}_T$, with a certificate (\hat{w}).
 - 3: **else**
 - 4: $w_d \notin \mathcal{B}_T$, with a certificate $w_{d,\mathcal{I}_g} \notin \mathcal{B}_{\mathcal{I}_g}$.
 - 5: **end if**
-

4.2 Using input/state/output representation

In the general case of a possibly uncontrollable system \mathcal{B} , we solve Problem 1 by using an input/state/output representation $\mathcal{B}(A,B,C,D)$ of \mathcal{B} . The solution is again a solvability test of a system of linear equations, where the coefficients matrix

$$P_T := \begin{bmatrix} 0 & I \\ \mathcal{O}_T(A,C) & \mathcal{I}_T(H) \end{bmatrix}$$

is block structured with blocks the extended observability matrix

$$\mathcal{O}_T := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{bmatrix}$$

and a low-triangular block-Toeplitz matrix $\mathcal{I}_T(H)$, constructed from the Markov parameters

$$H_0 := D \quad \text{and} \quad H_t := CA^{t-1}B, \quad \text{for } t \geq 1.$$

Proposition 5. Let \mathcal{B} be a controllable linear time-invariant system with an input/state/output representation $\mathcal{B} = \mathcal{B}(A,B,C,D)$. The sequence $w_d \in (\mathbb{R}_e^q)^T$ with missing values $w_{d,\mathcal{I}_m} = \text{NaN}$ and given values w_{d,\mathcal{I}_g} is a trajectory of \mathcal{B} if and only if the system of linear equations

$$w_{d,\mathcal{I}_g} = P_{\mathcal{I}_g}(M)v' \quad (v')$$

has a solution \hat{v}' , in which case a completion of the missing values is given by

$$\hat{w}_{\mathcal{I}_m} = P_{\mathcal{I}_m}\hat{v}'. \quad (\hat{w}')$$

Proof. For the finite T -samples long trajectory w of \mathcal{B} , we have

$$w = \begin{bmatrix} u \\ y \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & I \\ \mathcal{O}_T(A,C) & \mathcal{I}_T(H) \end{bmatrix}}_{P_T} \underbrace{\begin{bmatrix} x_{\text{ini}} \\ u \end{bmatrix}}_{v'} = P_T v'.$$

Therefore, $\mathcal{B}_T = \text{image}(P_T)$ and \hat{w}' is a trajectory of \mathcal{B} . Since, by (v'), \hat{w}' interpolates the given data w_{d,\mathcal{I}_g} , \hat{w}' is a solution of Problem 1. \square

The solution method presented in Proposition 5 is summarized in Algorithm 2.

Algorithm 2 Missing values completion with a model specified by an input/state/output representation.

Input: Matrices A, B, C, D defining an input/state/output representation $\mathcal{B} = \mathcal{B}(A, B, C, D)$ of the system and a sequence $w_d \in (\mathbb{R}_c^q)^T$.

- 1: **if** (v') has solution **then**
 - 2: $w_d \in \mathcal{B}_T$, with a certificate (\hat{w}') .
 - 3: **else**
 - 4: $w_d \notin \mathcal{B}_T$, with a certificate $w_{d, \mathcal{I}_g} \notin \mathcal{B}_{\mathcal{I}_g}$.
 - 5: **end if**
-

Note 6 (Computational cost). Solving (v) and (v') by Gaussian elimination (LU factorization) without exploiting the structure of the matrices $\mathcal{F}_T(M)$ and P_T requires $O(T^2)$ operations.

Note 7 (Approximate solution). If $w_d \notin \mathcal{B}_T$, (v) and (v') have no solution. Computing an approximate solution, using the least-squares criterion, leads to a "smoothed sequence" \hat{w} , which simultaneously approximates the given data and completes the missing values.

Note 8 (Online filtering with missing data). Finding a recursive online algorithm for solving (v) and (v') approximately in the least-squares sense would be a generalization of the classical Kalman filter for missing data estimation.

4.3 Examples

As stated in the introduction and in Section 5, special cases of missing data estimation with a given model are the classical simulation and output tracking control problems. In this section, we apply Algorithm 2 on these problems and give an interpretation of the steps of the algorithm.

Simulation as missing data estimation

The classical simulation problem is defined as follows.

Problem 9 (Simulation). Given a system \mathcal{B} , a "past" trajectory w_p (specifying the initial condition), and a "future" input u_f , find the corresponding future output y_f , *i.e.*, find y_f , such that the trajectory $w_d = w_p \wedge (u_f, y_f)$ obtained by concatenation of the past trajectory w_p and the future trajectory $w_f = (u_f, y_f)$ is a valid trajectory of \mathcal{B} .

In the model based simulation problem, the given data is a trajectory w_d of the system \mathcal{B} , with missing value y_f . Therefore, in the case of a linear time-invariant system \mathcal{B} , we can use Algorithm 2 to solve the problem.

With missing value y_f , equation (v') simplifies to the following formula for estimation of the initial state x_{ini} from the past trajectory w_p and the model parameters

$$\mathcal{O}_\ell(A, C)x_{\text{ini}} = y_p - \mathcal{F}_\ell(H)u_p. \quad (x_{\text{ini}})$$

By the assumption that w_p specifies the initial condition and by assuming, in addition, that the state space representation $\mathcal{B}(A, B, C, D)$ is minimal, the system of equations (x_{ini}) has a unique solution.

Then the data completion equation (\hat{w}') reduces to the formula for the calculation of the output from the given input u_f and the initial condition x_{ini} obtained from (v')

$$y_f = \mathcal{O}_{\ell+1:T}(A, C)x_{\text{ini}} + \mathcal{F}_{\ell+1:T}(H) \begin{bmatrix} u_p \\ u_f \end{bmatrix}.$$

Therefore, in the case of missing data y_f (Problem 9), Algorithm 2 reduces to the classical method for simulation of a state space model:

1. estimate the initial condition, and
2. simulate the output from the estimated initial condition and the given input.

Tracking control as missing data estimation

Consider now the output tracking control problem defined as follows.

Problem 10 (Tracking control). Given a system \mathcal{B} , a "past" trajectory w_p (specifying the initial condition), and a "future" output y_f , find a future input u_f , such that $w_d = w_p \wedge (u_f, y_f)$ is a trajectory of \mathcal{B} , or assert that such an input does not exist.

Viewed as a missing data problem, in Problem 10 we are given a trajectory w_d of the system \mathcal{B} , with missing value u_f . As in the previous example, we solve the problem by Algorithms 2 and relate equations (v') and (\hat{w}') to steps in the classical solution methods.

With missing value u_f , equation (v') decouples into two independent equations:

1. initial state estimate (x_{ini}) , and
2. output feasibility condition

$$y_f - (\mathcal{O}_{\ell+1:T}(A, C)x_{\text{ini}} + \mathcal{T}_{\ell+1:T, 1:\ell}(H)u_p) = \mathcal{T}_{\ell+1:T, \ell+1:T}(H)u_f. \quad (u_f)$$

While (x_{ini}) by assumption has a solution, equation (u_f) may not have a solution, depending on the given system, initial condition and desired output trajectory. If (u_f) is solvable, the control signal u_f is given by (v') .

In the context of the output tracking control, it is relevant to relax the exact tracking condition to an approximate one, *e.g.*, optimal tracking in the ℓ_2 -norm sense. The ℓ_2 -optimal tracking control is computed by solving (u_f) approximately in the least squares sense. In this case, it is well known that the Riccati recursion is an efficient way to solve the problem. Finding similar efficient recursive solution for the general Problem 1 and in particular for the computations in Algorithm 2 is an open problem.

5 Exact system identification from trajectory with missing values

In this section, two solution approaches are presented for Problem 2—exact system identification from trajectory with missing values. The first one is based on the nuclear norm heuristic and requires convex optimization. The second one is based on ideas from system theory and subspace identification and uses only linear algebra operations, such as computation of a null space of a matrix and reduction of a nonminimal representation to a minimal one.

5.1 Hankel structured matrix completion using the nuclear norm heuristic

Link to rank deficient Hankel matrices

Both the nuclear norm minimization method and the subspace method are based on a connection between trajectories w of linear time-invariant systems with bounded complexity and rank deficient Hankel matrices

$$\mathcal{H}_L(w) := \begin{bmatrix} w(1) & w(2) & \cdots & w(T-L+1) \\ w(2) & w(3) & \cdots & w(T-L+2) \\ w(3) & w(4) & \cdots & w(T-L+3) \\ \vdots & \vdots & & \vdots \\ w(L) & w(L+1) & \cdots & w(T) \end{bmatrix}.$$

Lemma 11 ([14]). *Let p and ℓ be, respectively, the number of outputs and the lag of a linear time-invariant system \mathcal{B} . Then, for $L \geq \ell + 1$,*

$$w \in \mathcal{B} \iff \text{rank}(\mathcal{H}_L(w)) \leq mL + p\ell. \quad (*)$$

Using $(*)$, we rewrite Problem 2 as the following Hankel structured low-rank matrix completion problem:

$$\text{minimize over } \hat{w} \quad \text{rank}(\mathcal{H}_L(\hat{w})) \quad \text{subject to} \quad \hat{w}_{\mathcal{I}_g} = w_{d, \mathcal{I}_g}. \quad (\text{HMC})$$

Due to the rank constraint, (HMC) is a nonconvex problem. A convex relaxation approach [4, 9] is to replace the rank by the nuclear norm $\|\cdot\|_*$ (sum of the singular values):

$$\text{minimize over } \hat{w} \in (\mathbb{R}^q)^T \quad \|\mathcal{H}_L(\hat{w})\|_* \quad \text{subject to } \hat{w}_{\mathcal{I}_g} = w_{d,\mathcal{I}_g}. \quad (\text{NN})$$

Problem (NN) is a semidefinite programming problem

$$\begin{aligned} & \text{minimize over } \hat{w}, U, \text{ and } V \quad \text{trace}(U) + \text{trace}(V) \\ & \text{subject to } \begin{bmatrix} U & \mathcal{H}_L(\hat{w})^\top \\ \mathcal{H}_L(\hat{w}) & V \end{bmatrix} \succeq 0 \quad \text{and } \hat{w}_{\mathcal{I}_g} = w_{d,\mathcal{I}_g} \end{aligned}$$

and can be solved globally and efficiently by existing methods [1]. Using general purpose semidefinite programming solvers, however, the computational cost is cubic in the length T of the sequence. Efficient solvers that exploit the Hankel structure as well as efficient first order optimization methods are currently being developed, see, *e.g.*, [11].

Example: completing the sequence of the Fibonacci numbers

Consider the sequence of the first 8 Fibonacci numbers with missing 3rd and 6th element:

$$w_d = (1, 1, \text{NaN}, 3, 5, \text{NaN}, 13, 21). \quad (w_d)$$

The complete sequence \bar{w} is a trajectory of an autonomous linear time-invariant system with lag $\ell = 2$

$$\bar{\mathcal{B}} = \ker(\bar{R}(\sigma)), \quad \text{where } \bar{R}(z) = 1 + z - z^2.$$

Therefore, the Hankel matrix $\mathcal{H}_L(\bar{w})$ has rank at most 2. The nuclear norm minimization problem (NN) for the example is

$$\text{minimize over } \hat{w}(3) \text{ and } \hat{w}(6) \quad \left\| \begin{bmatrix} 1 & 1 & \hat{w}(3) & 3 & 5 \\ 1 & \hat{w}(3) & 3 & 5 & \hat{w}(6) \\ \hat{w}(3) & 3 & 5 & \hat{w}(6) & 13 \\ 3 & 5 & \hat{w}(6) & 13 & 21 \end{bmatrix} \right\|_*.$$

We solve it with CVX [5]:

```
cvx_begin sdp;
    variables wh3 wh6;
    minimize(norm_nuc([ 1 1 wh3 3 5 ;
                       1 wh3 3 5 wh6 ;
                       wh3 3 5 wh6 13 ;
                       3 5 wh6 13 21 ])));
cvx_end
```

The default solver SDPT3 obtains in 10 iterations the solution $\hat{w}(3) = 2$ and $\hat{w}(6) = 8$.

Example: lightly damped mechanical system

As a realistic engineering example, consider the lightly damped mechanical system with poles

$$z_{1,2} = 0.8889 \pm i0.4402, \quad z_{3,4} = 0.4500 \pm i0.8801, \quad z_{5,6} = 0.6368 \pm i0.7673.$$

The identification data w_d is a $T = 200$ samples long trajectory, of which 86 samples are missing in a periodic pattern, see Figure 1. Using SDPT3 to solve the nuclear norm minimization problem (NN) for this data takes 96 seconds on a laptop with 2.60GHz Intel i7 processor and 8G RAM. The solution is correct up to the default convergence tolerance of the optimization method.

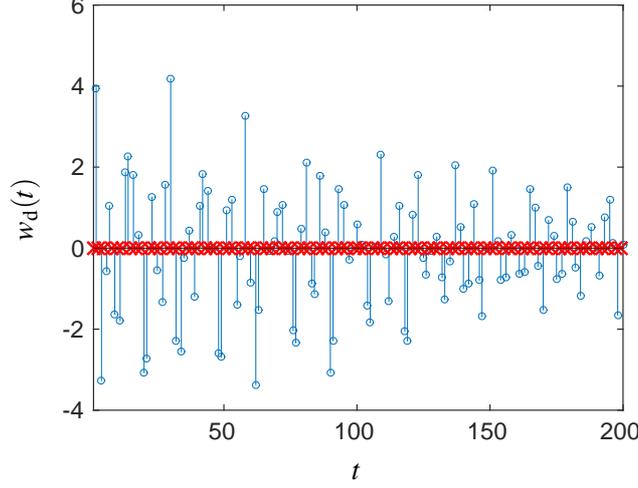


Figure 1: The data w_d for the identification of the lightly damped mechanical system. The red crosses denote the location of the missing samples.

5.2 Subspace method for identification with missing data

The method presented in this section is related to the class of the subspace identification methods [24]. The procedure is based on linear algebra operations and does not require nonlinear optimization. First, we illustrate the method on examples and then we present the general algorithm.

Example: autonomous system

Consider again the example of completing the sequence of the Fibonacci numbers (w_d). We select the the following complete submatrices

$$H^1 = \begin{bmatrix} 1 & 3 \\ 1 & 5 \\ 3 & 13 \end{bmatrix} \quad \text{and} \quad H^2 = \begin{bmatrix} 1 & 5 \\ 3 & 13 \\ 5 & 21 \end{bmatrix}$$

of the incomplete Hankel matrix

$$\mathcal{H}_4(w_d) = \begin{bmatrix} 1 & 1 & \text{NaN} & 3 & 5 \\ 1 & \text{NaN} & 3 & 5 & \text{NaN} \\ \text{NaN} & 3 & 5 & \text{NaN} & 13 \\ 3 & 5 & \text{NaN} & 13 & 21 \end{bmatrix}.$$

Since $\text{rank}(\mathcal{H}_4(\bar{w})) = 2$, H^1 and H^2 have nontrivial left null spaces. We have,

$$\begin{bmatrix} 1 & 2 & -1 \end{bmatrix} H^1 = 0 \quad \text{and} \quad \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} H^2 = 0.$$

With the convention $0 \times \text{NaN} = 0$, we extend the vectors in the left null spaces of H^1 and H^2 to vectors in the left null space of $\mathcal{H}_4(w_d)$ by adding zeros at the location of the missing values

$$\underbrace{\begin{bmatrix} 1 & 2 & 0 & -1 \\ 1 & 0 & -2 & 1 \end{bmatrix}}_{\tilde{R}} \mathcal{H}_4(w_d) = 0. \quad (**)$$

This shows that, the polynomial matrix

$$\tilde{R}(z) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} z^0 + \begin{bmatrix} 2 \\ 0 \end{bmatrix} z^1 + \begin{bmatrix} 0 \\ -2 \end{bmatrix} z^2 + \begin{bmatrix} -1 \\ 1 \end{bmatrix} z^3$$

obtained from (**) defines a kernel representation of $\mathcal{B}_{\text{mpum}}(w_d)$. Indeed the greatest common divisor of the first and second elements of $R(z)$ is a minimal kernel representation of the data generating system $\tilde{\mathcal{B}}$:

$$\hat{R}(z) := \text{GCD}(1 + 2z - z^3, 1 - 2z^2 + z^3) = 1 + z - z^2 = \bar{R}(z).$$

After the model is identified, the completion of w_d is an estimation problem that can be solved by Algorithms 1 or 2.

Example: open system

Consider the sequence with missing values

$$w_d = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \text{NaN} \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} \text{NaN} \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \begin{bmatrix} \text{NaN} \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ 10 \end{bmatrix}, \begin{bmatrix} \text{NaN} \\ 14 \end{bmatrix}, \begin{bmatrix} 5 \\ 9 \end{bmatrix} \right)$$

that is generated by a linear time invariant system

$$\tilde{\mathcal{B}} = \{ w \mid \underbrace{\begin{bmatrix} 1 & 1 \end{bmatrix}}_{\bar{R}_0} w(t) + \underbrace{\begin{bmatrix} 0 & -1 \end{bmatrix}}_{\bar{R}_1} w(t+1) = 0 \}$$

with $m = 1$ input and lag $\ell = 1$. The unknown complete sequence is

$$\bar{w} = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ 10 \end{bmatrix}, \begin{bmatrix} -5 \\ 14 \end{bmatrix}, \begin{bmatrix} 5 \\ 9 \end{bmatrix} \right) \in \tilde{\mathcal{B}}_T$$

Our task is to find $\tilde{\mathcal{B}}$ from the data w_d .

First, we revise the classical approach of finding $\tilde{\mathcal{B}}$ from a complete trajectory \bar{w} . The parameter vector

$$\bar{R} = [\bar{R}_0 \quad \bar{R}_1] = [1 \quad 1 \quad 0 \quad -1]$$

of a kernel representation $\ker(\bar{R}(\sigma))$ satisfies the equation

$$\bar{R} \mathcal{H}_2(\bar{w}) = 0.$$

Provided that the left null space of $\mathcal{H}_2(\bar{w})$ is one dimensional (persistence of excitation assumption [30], satisfied by \bar{w}), the kernel parameter \bar{R} can be found up to a scaling factor from \bar{w} by computing a basis for the left null space of $\mathcal{H}_2(\bar{w})$. The Hankel matrix $\mathcal{H}_2(w_d)$ of the given data

$$\mathcal{H}_2(w_d) = \begin{bmatrix} 1 & \text{NaN} & 2 & \text{NaN} & 3 & \text{NaN} & 4 & \text{NaN} \\ 0 & 1 & 1 & 3 & 2 & 5 & 10 & 14 \\ \text{NaN} & 2 & \text{NaN} & 3 & \text{NaN} & 4 & \text{NaN} & 5 \\ 1 & 1 & 3 & 2 & 5 & 10 & 14 & 9 \end{bmatrix},$$

however, has unspecified entries in every column, so that its left kernel can not be computed. Therefore, the classical method is not applicable.

The idea behind the method for exact identification with missing data is to consider the *extended Hankel matrix*

$$\mathcal{H}_3(w_d) = \begin{bmatrix} 1 & \text{NaN} & 2 & \text{NaN} & 3 & \text{NaN} & 4 \\ 0 & 1 & 1 & 3 & 2 & 5 & 10 \\ \text{NaN} & 2 & \text{NaN} & 3 & \text{NaN} & 4 & \text{NaN} \\ 1 & 1 & 3 & 2 & 5 & 10 & 14 \\ 2 & \text{NaN} & 3 & \text{NaN} & 4 & \text{NaN} & 5 \\ 1 & 3 & 2 & 5 & 10 & 14 & 9 \end{bmatrix}$$

and select the two submatrices of $\mathcal{H}_3(w_d)$

$$\tilde{H}^1 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 10 \\ \text{NaN} & \text{NaN} & \text{NaN} & \text{NaN} \\ 1 & 3 & 5 & 14 \\ 2 & 3 & 4 & 5 \\ 1 & 2 & 10 & 9 \end{bmatrix} \quad \text{and} \quad \tilde{H}^2 = \begin{bmatrix} \text{NaN} & \text{NaN} & \text{NaN} \\ 1 & 3 & 5 \\ 2 & 3 & 4 \\ 1 & 2 & 10 \\ \text{NaN} & \text{NaN} & \text{NaN} \\ 3 & 5 & 14 \end{bmatrix},$$

that have missing values in the same rows. The matrices H^1 and H^2 , obtained from \tilde{H}_1 and \tilde{H}_2 , respectively, by removing the rows of missing values have nontrivial left null spaces

$$\underbrace{\begin{bmatrix} -1 & -1 & 1 & 0 & 0 \end{bmatrix}}_{\tilde{R}^1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 10 \\ 1 & 3 & 5 & 14 \\ 2 & 3 & 4 & 5 \\ 1 & 2 & 10 & 9 \end{bmatrix} = 0 \quad \text{and} \quad \underbrace{\begin{bmatrix} 0 & -1 & -1 & 1 \end{bmatrix}}_{\tilde{R}^2} \begin{bmatrix} 1 & 3 & 5 \\ 2 & 3 & 4 \\ 1 & 2 & 10 \\ 3 & 5 & 14 \end{bmatrix} = 0.$$

Inserting zeros in the R^1 and R^2 at the location of the missing values, we obtain vectors \tilde{R}^1 and \tilde{R}^2 in the left null spaces of \tilde{H}^1 and \tilde{H}^2

$$\underbrace{\begin{bmatrix} -1 & -1 & 0 & 1 & 0 & 0 \end{bmatrix}}_{\tilde{R}^1} \tilde{H}^1 = 0 \quad \text{and} \quad \underbrace{\begin{bmatrix} 0 & 0 & -1 & -1 & 0 & 1 \end{bmatrix}}_{\tilde{R}^2} \tilde{H}^2 = 0.$$

By construction $\begin{bmatrix} \tilde{R}^1 \\ \tilde{R}^2 \end{bmatrix} \mathcal{H}_3(\tilde{w}) = 0$, so that, the polynomial matrix

$$\tilde{R}(z) = \begin{bmatrix} \tilde{R}^1(z) \\ \tilde{R}^2(z) \end{bmatrix} = \begin{bmatrix} [-1 & -1]z^0 + [0 & 1]z^1 + [0 & 0]z^2 \\ [0 & 0]z^0 + [-1 & -1]z^1 + [0 & 1]z^2 \end{bmatrix}$$

is a (nonminimal) null space representation of $\hat{\mathcal{B}}$. In the example, $\tilde{R}^2(z) = z\tilde{R}^1(z)$, so that a minimal null space representation is

$$\hat{R}(z) = [1 \quad 1] + [0 \quad -1]z = \bar{R}(z).$$

In general the reduction $\tilde{R}(z) \mapsto \hat{R}(z)$ of the identified nonminimal representation $\tilde{R}(z)$ to a minimal one $\hat{R}(z)$ requires computation, see [19]. Once the model is identified, filling in the missing data can be done by Algorithms 1 and 2. The computed complete sequence \hat{w} coincides with the true complete sequence \tilde{w} .

The subspace algorithm

The generalization of the procedure used in the examples is summarized in Algorithm 3.

Algorithm 3 Subspace algorithm for linear time-varying system identification with missing data.

Input: A sequence $w_d \in (\mathbb{R}_c^q)^T$ and natural numbers m .

1: Select real valued submatrices

$$H^i \in \mathbb{R}^{m_i \times n_i}, \quad \text{with } m_i \geq n_i - 1,$$

of the Hankel matrix $\mathcal{H}_T(w_{\text{ext}})$, where $w_{\text{ext}} = (w_d, \underbrace{\text{NaN}, \dots, \text{NaN}}_T)$.

2: Compute bases R^i for the left null spaces of H^i , i.e., full row rank matrices $R^i \in \mathbb{R}^{g_i \times m_i}$ of maximum row dimension g_i , such that $R^i H^i = 0$.

3: Extend $R^i \in \mathbb{R}^{g_i \times m_i}$ to $\tilde{R}^i \in \mathbb{R}^{g_i \times Tq}$ by inserting zero columns at the location of the rows removed from $\mathcal{H}_T(w_{\text{ext}})$ in the selection of H^i .

4: Compute a minimal kernel representation of \hat{R} of $\tilde{R}^1, \dots, \tilde{R}^K$, where K be the number of complete submatrices.

5: Convert $\hat{R}(z)$ to an input/state/output representation and apply Algorithm 2 to compute the complete sequence \hat{w} .

Output: $\hat{R}(z)$ and \hat{w} .

Algorithm 3 is implemented in Matlab at the code is available from:

<http://homepages.vub.ac.be/~imarkovs/software/mpum-md.tar>

Solving the identification problem with the lightly damped mechanical system (second example in Section 5.1) from the data in Figure 1, the exact system is recovered up to numerical precision errors, in 0.27 seconds, using the same laptop as in the other simulation example. This result demonstrates the lower computation cost of Algorithm 3 in comparison with the nuclear norm minimization method.

6 Conclusions and discussion

In this paper we have generalized the notion of the most powerful unfalsified model in the model class of linear time-invariant systems to time series with missing values. Another interpretation of the topic of the paper is interpolation and extrapolation of a vector sequence by a linear time-invariant system of smallest complexity.

First, we formulated and solved the simpler problem of completing a given time series with missing values when the data generating system is given. This problem is reminiscent to the state estimation problem in the classical linear system theory, and is called "estimation problem". The solution methods check solvability of a system of linear equations with structured coefficients matrix. Finding efficient $O(T)$, where T is the length of the time series, numerical methods is a topic of future research. This problem is also related to the generalization of the classical Kalman filter to estimation of missing data (see Note 8).

The methods proposed in the paper for the solution of the identification problem are based on an equivalent formulation of the problem as matrix completion with the constraint that the complete matrix is Hankel structured and low-rank. The first method is based on the nuclear norm heuristic and involves solution of a semidefinite programming problem. The second method is a subspace-type method and requires only standard linear algebra computations.

The two equivalent formulations of

1. exact linear time-invariant system identification from data with missing values,
2. low-rank Hankel matrix completion,

allow us to establish links between methods for matrix completion (nuclear norm heuristic) and methods in system theory and system identification (realization and subspace algorithms). On the one hand, the system theoretic view brings deeper understanding of the structure of rank deficient Hankel matrices as well as engineering applications of the more general matrix completion problem and solution methods. On the other hand, the matrix completion view of the estimation and system identification problems with missing values brings a new framework of posing and solving system theoretic problems. This framework makes possible to use new methods (most notably the convex relaxation methods based on the nuclear norm heuristic) for solving these problems.

Acknowledgments

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] S. Boyd and L. Vandenberghe. *Convex Optimization*. 2001.
- [2] A. Chiuso. "Asymptotic variance of closed-loop subspace identification methods". In: *IEEE Transactions on Automatic Control* 51.8 (2006), pp. 1299–1314. ISSN: 0018-9286.
- [3] C. Chou and M. Verhaegen. "Subspace algorithms for the identification of multivariate errors-in-variables models". In: *Automatica* 33.10 (1997), pp. 1857–1869.
- [4] M. Fazel. "Matrix rank minimization with applications". PhD thesis. Elec. Eng. Dept., Stanford University, 2002.
- [5] M. Grant and S. Boyd. *CVX: Matlab software for disciplined convex programming*. stanford.edu/~boyd/cvx.
- [6] Goethals I. et al. "Identification of Positive Real Models in Subspace Identification by using Regularization". In: *IEEE Trans. Automat. Contr.* 48 (2003), pp. 1843–1847.
- [7] R. E. Kalman. "On partial realizations, transfer functions, and canonical forms". In: *Acta Polytechnica Scandinavica* 31 (1979), pp. 9–32.
- [8] R. E. Kalman, P. L. Falb, and M. A. Arbib. *Topics in Mathematical System Theory*. McGraw-Hill, 1969.

- [9] Z. Liu and L. Vandenberghe. “Interior-point method for nuclear norm approximation with application to system identification”. In: *SIAM J. Matrix Anal. Appl.* 31.3 (2009), pp. 1235–1256.
- [10] L. Ljung. *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1999.
- [11] D. Sun P. Tseng M. Fazel T. K. Pong. “Hankel matrix rank minimization with applications in system identification and realization”. In: *SIAM J. Matrix Anal. Appl.* 34.3 (2013), pp. 946–977.
- [12] I. Markovsky. “A software package for system identification in the behavioral setting”. In: *Control Engineering Practice* 21 (10 2013), pp. 1422–1436. DOI: 10.1016/j.conengprac.2013.06.010.
- [13] I. Markovsky. “Exact identification with missing data”. In: *Proc. of the 52nd IEEE Conference on Decision and Control*. Florence, Italy, 2013, pp. 151–155. DOI: 10.1109/CDC.2013.6759874.
- [14] I. Markovsky. *Low Rank Approximation: Algorithms, Implementation, Applications*. Communications and Control Engineering. Springer, 2012. DOI: 10.1007/978-1-4471-2227-2.
- [15] I. Markovsky. “Recent progress on variable projection methods for structured low-rank approximation”. In: *Signal Processing* 96PB (2014), pp. 406–419. DOI: 10.1016/j.sigpro.2013.09.021.
- [16] I. Markovsky. “Structured low-rank approximation and its applications”. In: *Automatica* 44.4 (2008), pp. 891–909.
- [17] I. Markovsky and K. Usevich. “Structured low-rank approximation with missing data”. In: *SIAM J. Matrix Anal. Appl.* (2 2013), pp. 814–830. DOI: 10.1137/120883050.
- [18] I. Markovsky et al. *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*. Monographs on Mathematical Modeling and Computation 11. SIAM, 2006. DOI: 10.1137/1.9780898718263.
- [19] J. Polderman and J. C. Willems. *Introduction to Mathematical Systems Theory*. New York: Springer-Verlag, 1998.
- [20] P. Rapisarda and H. Trentelman. “Identification and data-driven model reduction of state-space representations of lossless and dissipative systems from noise-free data”. In: *Automatica* 47.8 (2011), pp. 1721–1728.
- [21] J. Schoukens et al. “Frequency Response Function Measurements Using Concatenated Subrecords With Arbitrary Length”. In: *IEEE Trans. on Instr. and Measurement* 61.10 (2012), pp. 2682–2688.
- [22] T. Söderström and P. Stoica. *System Identification*. Prentice Hall, 1989.
- [23] K. Usevich and I. Markovsky. “Variable projection for affinely structured low-rank approximation in weighted 2-norms”. In: *J. Comput. Appl. Math.* 272 (2014), pp. 430–448. DOI: 10.1016/j.cam.2013.04.034.
- [24] P. Van Overschee and B. De Moor. *Subspace identification for linear systems: Theory, implementation, applications*. Boston: Kluwer, 1996.
- [25] M. Verhaegen and P. Dewilde. “Subspace model identification, Part 1: The output-error state-space model identification class of algorithms”. In: *Int. J. Control* 56 (1992), pp. 1187–1210.
- [26] R. Wallin and A. Hansson. “Maximum likelihood estimation of linear SISO models subject to missing output data and missing input data”. In: *Int. J. Control* 87.11 (2014), pp. 2354–2364.
- [27] J. C. Willems. “From time series to linear system—Part I. Finite dimensional linear time invariant systems”. In: *Automatica* 22.5 (1986), pp. 561–580.
- [28] J. C. Willems. “From time series to linear system—Part II. Exact modelling”. In: *Automatica* 22.6 (1986), pp. 675–694.
- [29] J. C. Willems. “From time series to linear system—Part III. Approximate modelling”. In: *Automatica* 23.1 (1987), pp. 87–115.
- [30] J. C. Willems et al. “A note on persistency of excitation”. In: *Control Lett.* 54.4 (2005), pp. 325–329. DOI: 10.1016/j.sysconle.2004.09.003.