Partial-realization theory and algorithms for linear switched systems A formal power series approach

Mihály Petreczky

Maastricht University P.O. Box 616, 6200 MD Maastricht, The Netherlands
M.Petreczky@maastrichtuniversity.nl

Jan H. van Schuppen

Centrum voor Wiskunde en Informatica (CWI), P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
J.H.van.Schuppen@cwi.nl

Abstract

The paper presents partial-realization theory and realization algorithms for linear switched systems. Linear switched systems are a particular subclass of hybrid systems. We formulate a notion of a partial realization and we present conditions for existence of a minimal partial realization. We propose two partial-realization algorithms and we show that under certain conditions they yield a complete realization. Our main tool is the theory of rational formal power series.

Key words: Hybrid systems, linear switched systems,, partial-realization theory, realization algorithm, formal power series
1991 MSC: 93B15, 93B20, 93B25, 93C99

1 Introduction

The main objective of the paper
The immediate objective of the paper is to present partial-realization theory of linear switched systems. However, the broader goal is to demonstrate that

(a) it is possible to develop partial-realization theory for hybrid systems,
(b) the theory of rational formal power series can be used to obtain partial-realization theory of hybrid systems.

The results of the paper just serve as evidence for the above claims. In fact, partial-realization theory can be developed along the same line for other classes of hybrid systems, [44]. This paper is intended as the first one in a series of papers dealing with partial-realization theory of hybrid systems.

The class of linear switched systems

Linear switched systems have been studied for almost two decades, see [35,51] for a survey. Their practical and theoretical relevance is widely recognized. *Linear switched systems (abbreviated by LSS)* are hybrid systems with external switching only, whose continuous dynamics in each discrete mode is determined by a linear continuous-time system, and whose discrete dynamics is trivial, i.e. any discrete state transition is allowed. The continuous subsystems are defined on the same state-, input- and output-spaces. A discrete state transition occurs if the environment enforces one, i.e. the switching sequence itself functions as an input. The time evolution of a linear switched system can be pictured as follows. A discrete mode is selected and a continuous input is fed into the corresponding linear subsystem. The state and output of the linear switched system is now determined by the time evolution of the chosen linear subsystem. At some point in time, a new discrete mode is selected. Then the linear subsystem corresponding to the new discrete mode is launched, using as initial state the state of the previous linear subsystem at the moment when a transition to the new mode occurred. From this point on, the state and output evolve according to the newly selected linear subsystem.

Motivation

We believe that similarly to linear systems, partial-realization theory for LSSs will be useful for systems identification and model reduction of LSSs and even of more general hybrid systems. We will elaborate on the possible applications in Section 8.

Contribution of the paper

The paper presents the following results.

- **Partial-realization algorithm and theorem for LSSs**
  In Section 6 we propose a notion of partial-realization. In Theorem 3 and 4 we formulate a sufficient condition for existence and minimality of a partial realization. We present two algorithms for computing a partial-realization. In one of them, Algorithm 1, a partial-realization is constructed from the columns of a *finite* sub-matrix Hankel-matrix, the other one, Algorithm 2, is based on finding a factorization of a *finite* sub-matrix of the Hankel-matrix. The outcomes of both procedures are isomorphic. The former is potentially useful for theoretical purposes, while the latter one might serve as a basis for subspace identification-like methods. The factorization algorithm was implemented and the implementation is available from the the first-named author on request.

- **Realization algorithm for LSSs**
  In Theorem 3 and 5 we show that the above partial-realization algorithms return a complete minimal realization of the input-output behavior, provided the rank of the finite sub-matrix of the Hankel-matrix equals the rank of the full Hankel-matrix. We show that any finite sub-matrix whose size is above a certain lower bound has this property.
- **Partial-realization theory for families of formal power series**
  Both the (partial-) realization algorithm and the corresponding theoretical results for LSSs follow quite easily from analogous statements for families of rational formal series. In Appendix B we present what can be termed as partial-realization theory and algorithms for *families of formal power series*. Although partial-realization theory of a single rational formal power series has been known in various forms [29,15,31,49,50,11,7], its extension to families of formal power series appears to be new. As it was already noted, this extension of the classical theory can be then used to develop partial-realization theory for a number of other (more general) classes of hybrid systems, [44].

Formal power series approach
Our main tool is the theory of rational formal power series. Recall from [42,45,40] that there is a correspondence between LSS realizations and representations of certain rational families of former power series. Recall that the Hankel-matrix of a family of input-output maps is in fact the Hankel-matrix of the corresponding family of formal power series. Hence, by formulating a partial-realization theory and algorithms for rational families of formal power series, we immediately obtain partial-realization theory and algorithms for LSSs.

Related work
To the best of our knowledge, the only results on partial-realization of linear switched systems is the first-named author's thesis [44]. In [40,39] partial-realization theory of bilinear hybrid and switched systems are announced, but no proofs are provided. The thesis [44] contains most of the results of this paper. In addition, it also covers partial-realization theory of several other classes of hybrid systems. Reference [47] announces some results on partial-realization theory of stochastic jump-Markov systems.

Realization theory of rational formal power series is a classical topic, see [49,50,30,2,34,12,13]. It is known to be closely related to realization theory of bilinear and state-affine systems [29,31,49,50]. In turn, results on partial-realization theory for discrete-time bilinear systems and state-affine systems can be found in [15,29,49,50,11,7]. With respect to [29,15,31,49,50,11,7], the main novelty of the paper is the following.

(1) In this paper (in Appendix B) partial-realization theory is stated directly

for rational formal power series representations, without reformulating it in terms of some particular system class such as bilinear or state-affine systems.

(2) While the classical results [29,15,31,49,50,11,7] can be thought of as a reformulation of partial-realization theory for a single rational formal power series, here we deal with *families of formal power series*.

(3) We apply partial-realization theory of rational formal power series to linear switched systems. This represents a novel result in the theory of hybrid systems. We view this as the main contribution of the paper.

The statement of the main results on partial-realization theory for families of formal power series was announced [43,47,39,40,41,46], but no proof was ever presented. The thesis [44] contains the statement and the proof of the main results on formal power series.

Outline of the paper

The outline of the paper is the following. Section 2 presents an informal formulation of the partial-realization problem for LSSs. Section 3 discusses a numerical example. It is intended as an accessible explanation of the main results of the paper by means of an example. Section 4 presents the definition and some elementary properties of linear switched systems together with the basic notation and terminology. Section 5 provides a brief overview of the realization theory of linear switched systems. This section is a prerequisite for understanding the main results of the paper. Section 6 presents the formal statement of the main results of the paper. Section 7 presents the proof of the results stated in Section 6. Finally, in Section 8 we formulate the conclusions of the paper and we discuss potential applications of the presented results. As it was already remarked, partial-realization theory of LSSs is based on partial-realization theory of rational formal power series. Therefore, in Appendix A we included a brief overview of the extension of the theory of formal power series to families of formal power series. More precisely, Appendix A reviews the relevant concepts and results on formal power series from [42,40,45]. In Appendix B we present partial-realization theory of families of formal power series. and algorithms for computing a representation for a family of formal power series. Appendix B is in fact the main technical tool of the paper. It is a prerequisite for understanding the proof of the main results on partial-realization theory of LSSs.

2 Informal problem formulation

The goal of this section is to state the partial-realization problem for linear switched systems in an informal way. The next section, Section 3, provides a numerical example to illustrate the problem and the solution. We defer the formal problem statement until Section 6.

Brief review of linear-partial realization theory

The partial-realization problem was originally formulated for linear systems. [33,28,52] as follows. Assume that the first $N \in \mathbb{N}$ Markov parameters $S = \{M_i\}_{i=1}^N$ of an input-output map f are specified.

(a) Find conditions for existence of a linear system whose first N Markov-parameters coincide with $S = \{M_i\}_{i=1}^N$. Such a linear system is called a *partial-realization of* S. Characterize minimal partial-realizations. Find an algorithm for computing a (minimal) partial-realization.
(b) Find conditions under which the thus obtained linear system is a minimal realization of the input-output map f.

Notice that for discrete-time linear systems, the kth Markov parameter coincides with the output of the system at time k for a particular input. Hence, the partial-realization problem can be seen as an identification problem. In fact, partial-realization theory of linear systems can be used for systems identification and model reduction.

Markov-parameters for LSSs

Let Φ be a family of input-output map which map continuous-valued inputs and switching sequences to continuous outputs. That is, the elements of Φ are of the same form as the input-output maps generated by LSSs. Below we will define *generalized Markov-parameters* of a family Φ of input-output maps which could potentially be realized by a LSS. The generalized Markov-parameters of Φ are defined as certain high-order derivatives with respect to the switching times of the elements of Φ. More precisely, the Markov-parameters are indexed by triples consisting of the following components

(a) elements of Φ, or pairs (q, j) where q runs through the set of discrete modes and j runs through the set $\{1, 2, \ldots, m\}$, where m denotes the number of continuous-valued input channels, and
(b) sequences of discrete modes, and
(c) discrete modes.

Without going into details, the intuition behind the presented indexing is the following. A Markov-parameter of Φ indexed by $f \in \Phi$, a discrete mode q, and a sequence of discrete modes $q_1, q_2, \ldots, q_k$, stands for the partial derivative of f with respect to the switching times, evaluated at zero, for the following switching scenario. The system goes through the modes $q_1, q_2, \cdots, q_k$ and then jumps to q. Here the continuous input is set to zero. The intuitive meaning of a Markov-parameter indexed by a pair (q_0, j), a discrete mode q and a sequence of discrete modes $q_1, q_2, \ldots, q_k$ is the following. Each input-output map $f \in \Phi$ can be written as a sum $f = a_f + y^{\Phi}$ of two maps, where a_f is independent of the continuous input and y^{Φ} is common for all the elements of Φ and it is linear in continuous inputs. Roughly speaking, a_f accounts for the output from a certain initial state under zero input, and y^Φ represents the input-output map induced by the zero initial state. The Markov-parameter indexed by the pair (q_0, j), discrete mode q and sequence $q_1, q_2, \cdots q_k$ is the derivative of y^Φ with respect to the switching times evaluated at zero, for the switching scenario where the system goes from mode q_0 to $q_1, q_2, \cdots q_k$ and ends in q, and all the continuous input channels are 0 except the jth one which is 1.

Partial-realization problem for LSSs
We will refer to a Markov parameter indexed by a sequence of discrete modes of length k as a Markov-parameter of *order k*. Fix a natural number $N > 0$ and let S be the collection of all Markov-parameters of Φ of order at most N. That is, S is simply a collection of high-order derivatives of the elements of Φ, such that the degree of derivation is bounded by N. A LSS Σ is said to be a *N-partial realization* of Φ, if certain products of the matrices of Σ are equal to the corresponding elements of S. In other words, Σ is a N-partial realization of Φ if the input-output maps of Σ and those of Φ have the property that their derivatives corresponding to the Markov-parameters of order at most N coincide. From realization theory of LSSs [44,45,42] it follows that a LSS is a realization of Φ if and only if it is a N-partial realization of Φ for all $N \in \mathbb{N}$. The *partial realization problem* for LSSs can be now stated as follows.

- Find conditions for existence of a N-partial realization of Φ by LSS. Characterize minimal dimensional N-partial realizations of Φ. Find an algorithm for computing a minimal N-partial LSS realizations of Φ.
- Find conditions under which a minimal N-partial LSS realization of Φ is a complete realization of Φ.

The motivation for studying the partial-realization problem for LSSs is similar to that of for linear systems, i.e. we expect it to be useful for model reduction and systems identification. In Section 8 we will present a more detailed description of the motivation and possible applications.

3 Numerical example

The purpose of this section is to demonstrate the main results of the paper by means of a numerical example. In this section we will tacitly use the notation and terminology of Section 4 and Section 6.

Consider the linear switched system of the form

$$\Sigma \begin{cases} \dot{x}(t) = A_{q(t)}x(t) + B_{q(t)}u(t) \\ y(t) = C_{q(t)}x(t) \end{cases} \tag{1}$$

where $q(t) \in \{q_1, q_2\}$ is the discrete mode at time t, $x(t) \in \mathbb{R}^5$ is the continuous-state at t, $y(t) \in \mathbb{R}$ is the scalar output at t, and $u(t) \in \mathbb{R}$ is the scalar input at t. The system matrices A_q, B_q, C_q describing the linear control system residing in a state (mode) $q \in \{q_1, q_2\} = Q$ are of the following form.

$$A_{q_1} = \begin{bmatrix} 0\;0\;0\;0\;0 \\ 0\;0\;1\;0\;0 \\ 0\;0\;0\;1\;0 \\ 0\;0\;0\;0\;1 \\ 0\;0\;0\;0\;0 \end{bmatrix}, B_{q_1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, C_{q_1} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T$$

$$A_{q_2} = \begin{bmatrix} 0\;0\;0\;0\;0 \\ 0\;2\;0\;0\;0 \\ 0\;0\;0\;0\;0 \\ 0\;0\;0\;0\;0 \\ 0\;0\;0\;0\;3 \end{bmatrix}, B_{q_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, C_{q_2} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}^T$$

Consider the initial states $x_1 = 0$ and $x_2 = \begin{bmatrix} 0\;0\;0\;0\;1 \end{bmatrix}^T$. Consider the set of input-output maps $\Phi = \{f_1, f_2\}$ such that f_i is realized by Σ from the initial state x_i, $i = 1, 2$, i.e. $f_i(u, w) = y_\Sigma(x_i, u, w)$, for $i = 1, 2$, for each continuous-valued input $u \in PC(T, \mathcal{U})$ and finite switching sequence $w \in (Q \times T)^+$.

Notice that Φ can be realized by the following minimal linear switched system

$$\Sigma_m \begin{cases} \dot{x}(t) = A^m_{q(t)} x(t) + B^m_{q(t)} u(t) \\ y(t) = C^m_{q(t)} x(t) \end{cases} \tag{2}$$

where for each $q \in \{q_1, q_2\}$, the matrices A^m_q, B^m_q, C^m_q are of the following form

$$A^m_{q_1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, B^m_{q_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, C^m_{q_1} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}^T$$

$$A^m_{q_2} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, B^m_{q_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, C^m_{q_2} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T$$

More precisely, Σ_m realizes f_i from the initial states x^m_i, for $i = 1, 2$. Here, $x^m_1 = 0$ and $x^m_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^T$.

Since Φ has a realization by a LSS, it is clear that it has a generalized kernel representation and hence its Markov-parameters can be defined. In order to give a better intuition on Markov-parameters of Φ, we have listed some of them in Table 1. Consider the upper-left sub-matrix $H_{\Phi,K,L}$ of the Hankel-matrix H_Φ of Φ. Recall that $H_{\Phi,K,L}$ is formed by the intersection of the columns of H_Φ indexed by a sequence of discrete modes of length at most L, and by the rows of H_Φ indexed by a sequence of discrete modes of length at most K. From Theorem 3 it follows that rank $H_{\Phi,K,L} \leq \dim \Sigma_m = 4$ for all $K, L \geq 4$. In fact, it turns out that the Hankel-matrix $H_{\Phi,N,N}$ for $N = 2$ has already rank 4. By Theorem 3 it means that we can already compute a minimal LSS realization of Φ from the generalized Markov parameters of Φ of indexed by sequences of discrete modes of length at most $5 = 2 + 3$. Applying Algorithm 2 to $H_{\Phi,3,2}$ yields the following minimal LSS realization Σ_f of Φ.

$$\Sigma_f : \begin{cases} \dot{x}(t) = A^f_{q(t)} x(t) + B^f_{q(t)} u(t) \\ y(t) = C^f_{q(t)} x(t) \end{cases} \quad (3)$$

Table 1
Markov-parameters of Φ

$(q_0, j) \in (\{q_1, q_2\} \times \{1\})$	$q \in Q$	$w \in Q^*$	Markov-parameter $S_{q,q_0,j}(w)$
$(q_2, 1)$	q_2	ϵ	0
$(q_2, 1)$	q_2	q_1	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(q_2, 1)$	q_2	$q_2q_2q_2$	0
$(q_2, 1)$	q_1	ϵ	1
$(q_2, 1)$	q_1	q_1	0
$(q_2, 1)$	q_1	q_2	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(q_2, 1)$	q_1	q_2q_2	4
$(q_2, 1)$	q_1	$q_2q_2q_2$	8
$(q_1, 1)$	q_1	ϵ	0
$(q_1, 1)$	q_1	q_1	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(q_1, 1)$	q_1	$q_2q_2q_2$	0
$(q_1, 1)$	q_2	ϵ	1
$(q_1, 1)$	q_2	q_1	0
$(q_1, 1)$	q_2	q_2	3
$(q_1, 1)$	q_2	q_2q_2	9
$(q_1, 1)$	q_2	$q_2q_2q_2$	27
$j \in \Phi$	$q \in Q$	$w \in Q^*$	Markov-parameter $S_{j,q}(w)$
f_1	q_1	ϵ	0
f_1	q_1	q_1	0
f_2	q_1	ϵ	0
f_2	q_1	q_1	0
f_2	q_2	ϵ	1
f_2	q_2	q_1	0
f_2	q_2	q_2	3
f_2	q_2	q_2q_2	9
f_2	q_2	$q_2q_2q_2$	27

where the system matrices are of the form

$$A^f_{q_1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.79 \\ -0.49 & 0 & 0 & 0 \\ 0 & 0 & -2.55 & 0 \end{bmatrix}, B^f_{q_1} = \begin{bmatrix} 1.46 \\ 0 \\ 0 \\ 0 \end{bmatrix}, C^f_{q_1} = \begin{bmatrix} 0 \\ 0.71 \\ 0 \\ 0 \end{bmatrix}^T$$

$$A^f_{q_2} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, B^f_{q_2} = \begin{bmatrix} 0 \\ 1.42 \\ 0 \\ 0 \end{bmatrix}, C^f_{q_2} = \begin{bmatrix} 0.69 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The LSS Σ_f realizes the input-output map f_1 from the initial state $x^f_1 = (0,0,0,0)^T$, and the map f_2 from the initial state $x^f_2 = (1.46,0,0,0)^T$.

However, it turns out that rank $H_{\Phi,1,0}$ = rank $H_{\Phi,0,1}$ = rank $H_{\Phi,0,0}$. That is, the sequence of Markov parameters of Φ indexed by sequences of discrete modes of length at most 1 already satisfies the partial realization theorems Theorem 3 and Theorem 5. Hence, we can apply Algorithm 2 to $H_{\Phi,1,0}$ to obtain a minimal LSS 1-partial realization of Φ. However, it can be checked that the thus obtained LSS *is not a LSS realization of* Φ. Indeed, by applying Algorithm 2 to $H_{\Phi,1,0}$ we obtain the following LSS realization of Φ

$$\Sigma_{part} : \begin{cases} \dot{x}(t) = A^{part}_{q(t)} x(t) + B^{part}_{q(t)} u(t) \\ y(t) = C^{part}_{q(t)} x(t) \end{cases} \tag{4}$$

where the system matrices $A^{part}_q, B^{part}_q, C^{part}_q$, $q \in \{q_1, q_2\} = Q$ are as follows

$$A^{part}_{q_1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, B^{part}_{q_1} = \begin{bmatrix} -1.5 \\ 0 \end{bmatrix}, C^{part}_{q_1} = \begin{bmatrix} 0 \\ 0.67 \end{bmatrix}^T$$

$$A^{part}_{q_2} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}, B^{part}_{q_2} = \begin{bmatrix} 0 \\ 1.5 \end{bmatrix}, C^{part}_{q_2} = \begin{bmatrix} -0.67 & 0 \end{bmatrix}$$

The LSS Σ_{part} is a *partial realization* of the input-output map f_1 from the initial state $x^{part}_1 = (0,0)^T$, and of the input-output map f_2 from the initial state $x^{part}_2 = (-1,5,0)^T$. That is, the Markov parameters of $\Phi = \{f_1, f_2\}$ which are indexed by sequences from the set $\{\epsilon, q_1, q_2\}$ coincide with those of the input-output maps induced by the initial state x^{part}_i, $i = 1,2$. It is easy to see that Σ_{part} is not a realization of the input-output maps f_i from the respective initial states x^{part}_i, $= 1,2$. That is, the input-output maps induced by the respective states x_i^{part}, $i = 1, 2$, do not coincide with the maps f_i. One can either check it by direct calculation, or by using uniqueness of a minimal realization. Using the latter approach, it is enough to notice that Σ_m is a minimal realization of Φ and it is of dimension 4. Since Σ_{part} of dimension 2, hence smaller than the dimension of Σ_m, and all minimal realizations have to be of the same dimension, it follows that Σ_{part} cannot be a realization of Φ.

In fact, by checking the Markov-parameters, one can see that Σ_{part} recreates only the Markov-parameters indexed by sequences of length at most 1, but there is a Markov-parameter of Φ, indexed by a sequence of length 2, which is not generated by Σ_{part}. That is, it indeed happens that a family of input-output maps generated by a LSS satisfies the sufficient conditions for existence of a N-partial realization for some N, but the obtained partial realization is not a complete realization of the family of input-output maps.

4 Linear switched systems

This section contains the definition of linear switched systems. We will start with fixing notation and terminology which will be used throughout the paper. The notation used in this paper is mostly the standard one used in the field of control theory and formal language theory. In order to make the task of the reader easier, below we will list the most important notational conventions, grouped according to the disciplines.

4.0.1 Notation from general mathematics and control theory

Denote by T the set $[0, +\infty) \subseteq \mathbb{R}$ of all non-negative reals. The set T will be the time-axis of the systems discussed in this paper. For any $m \geq 0$, denote by $PC(T, \mathbb{R}^m)$ the class of piecewise-continuous maps from T to $\mathbb{R}^m$. That is, $f \in PC(T, \mathbb{R}^m)$, if f has finitely many points of discontinuity on each finite interval $[0, t]$, $t \in T$, and at each point of discontinuity the right- and left-hand side limits exist and they are finite. Denote by $\mathbb{N}$ the set of natural number including 0. By abuse of notation we will denote any constant function $f : T \to \mathbb{R}^m$ by its value. That is, if $f(t) = a \in \mathbb{R}^m$ for all $t \in T$, then f will be denoted by a. For any function g the range of g will be denoted by $\text{Im}g$, i.e. if $g : A \to B$ for some sets A and B, then $\text{Im}g = \{g(a) \in B \mid a \in A\}$. If $\mathcal{X}$ is a vector space and Z is a subset of $\mathcal{X}$, then $\text{Span}Z$ denotes the linear span of elements of Z in $\mathcal{X}$. If $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are vector spaces over $\mathbb{R}$, and $F_1 : \mathcal{X} \to \mathcal{Y}$, $F_2 : \mathcal{Y} \to \mathcal{Z}$ are linear maps, then F_1F_2 denotes the composition F_1 and F_2.

Let $\phi : \mathbb{R}^k \to \mathbb{R}^{p \times m}$ be a smooth map. Consider a k tuple of natural numbers $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k) \in \mathbb{N}^k$, where $\alpha_1, \ldots, \alpha_k \in \mathbb{N}$. We will denote $D^\alpha \phi$ the partial derivative of $\phi(t_1, t_2, \ldots, t_k)$ evaluated at zero, such that the order of derivation with respect to the variable t_i is α_i for $i = 1, \ldots, k$. That is,

$$D^\alpha \phi = \frac{d^{\alpha_1}}{dt_1^{\alpha_1}} \frac{d^{\alpha_2}}{dt_2^{\alpha_2}} \cdots \frac{d^{\alpha_k}}{dt_k^{\alpha_k}} \phi(t_1, t_2, \ldots, t_k)|_{t_1=t_2=\cdots=t_k=0}.$$

For each $i = 1, 2, \ldots, n$, e_j denotes the jth unit vector of $\mathbb{R}^n$, i.e. $e_j = (\delta_{1,j}, \delta_{2,j}, \ldots, \delta_{n,j})$, where $\delta_{i,j}$ is the Kronecker symbol.

4.0.2 Infinite matrices

In this paper we will use the notation of [32] for matrices indexed by sets other than natural numbers. Let I and J be two arbitrary sets. A (real) matrix M whose columns are indexed by the elements of J and whose rows are indexed by the elements of I is simply a map $M : I \times J \to \mathbb{R}$. The set of all such matrices is denoted by $\mathbb{R}^{I\times J}$. The entry of M indexed by the row index $i \in I$ and column index $j \in J$ is denoted by $M_{i,j}$ and it is defined as the value of M at (i, j), i.e. $M_{i,j} = M(i, j)$. The case of usual finite matrices can be recovered by viewing $n \times m$ real matrices as matrices from $\mathbb{R}^{\{1,2,\ldots,n\}\times\{1,2,\ldots,m\}}$. In the sequel, when referring to the index set of a matrix, we will identify any natural number n with the set $\{1, 2, \ldots, n\}$. In other words, $\mathbb{R}^{I\times n}$ denotes the set of matrices $\mathbb{R}^{I\times\{1,2,\ldots,n\}}$ and $\mathbb{R}^{n\times J}$ denotes the set of matrices $\mathbb{R}^{\{1,2,\ldots,n\}\times J}$.

For a matrix $M \in \mathbb{R}^{I\times J}$, the columns of M are simply maps of the form $I \to \mathbb{R}$ and the rows of M are maps of the form $J \to \mathbb{R}$. The set of maps of the form $I \to \mathbb{R}$ and $J \to \mathbb{R}$ will sometimes be denoted by $\mathbb{R}^I$ and $\mathbb{R}^J$ respectively. Furthermore, if $g \in \mathbb{R}^I$ (resp. $g \in \mathbb{R}^J$) then the value of g at $i \in I$ (resp. $j \in J$) will be denoted by g_i (resp. g_j). The column of M indexed by $j \in J$ will be denoted by $M_{.,j}$ and is defined as $M_{.,j}(i) = M_{i,j}$, $i \in I$. Similarly, the row if M indexed by $i \in I$ will be denoted by $M_{i,.}$ and is defined as $M_{i,.}(j) = M_{i,j}$ for all $j \in J$. If $M \in \mathbb{R}^{I\times J}$ and $S \in \mathbb{R}^{J\times K}$ and J is finite, then the product of M and S is the matrix $MS \in \mathbb{R}^{I\times K}$ such that $(MS)_{i,k} = \sum_{j\in J} M_{i,j} S_{j,k}$ for all $i \in I, k \in K$. In particular, if $M \in \mathbb{R}^{I\times r}$ and $S \in \mathbb{R}^{r\times K}$ for some natural number $r \in \mathbb{N}$, then their product MS is well-defined and it belongs to $\mathbb{R}^{I\times K}$.

We will identify a map $f \in \mathbb{R}^J$ with the matrix $f \in \mathbb{R}^{J\times 1}$ defined as $f_{j,1} = f_j$ for all $j \in J$. Hence, for a matrix $M \in \mathbb{R}^{I\times J}$, the product Mf is defined as the following matrix in $\mathbb{R}^{I\times 1}$; $(Mf)_{i,1} = \sum_{j\in J} M_{i,j} f_j$. In addition, we will occasionally identify the rows of a matrix $M \in \mathbb{R}^{I\times J}$ with matrices $\mathbb{R}^{1\times J}$. That is, the row $M_{i,.}$ of M indexed by $i \in I$ will be viewed as the matrix $M_{i,.} : \{1\} \times J \ni (1, j) \mapsto M_{i,j}$. With this identification, the product of the row $M_{i,.}$ with f is a scalar $M_{i,.} f = \sum_{j\in J} M_{i,j} f_j \in \mathbb{R}$. Notice that here we tacitly assumed that J is finite.

Notice that the set of all maps $\mathbb{R}^I$ forms a vector space with respect to pointwise addition and multiplication by scalar. That is, if $f, g \in \mathbb{R}^I$ and $\alpha, \beta \in \mathbb{R}$, then the linear combination $\alpha f + \beta g : I \to \mathbb{R}$ is defined by $(\alpha f + \beta g)(i) = \alpha f(i) + \beta g(i)$ for all $i \in I$. Consider a matrix $M \in \mathbb{R}^{I \times J}$ and recall that its columns are simply elements of $\mathbb{R}^I$. Hence, it makes sense to speak of the linear subspace spanned by the columns of a matrix $M \in \mathbb{R}^{I \times J}$. In the sequel, the *rank of* M, denoted by rank $M \in \mathbb{N} \cup \{\infty\}$ will mean the dimension of the linear space spanned by the columns of M. If this dimension is not finite, then the rank is taken to be ∞. We will denote by $\text{Im}M$ the linear space spanned by the columns of M.

If $M \in \mathbb{R}^{I \times J}$ and J is finite, then M can be viewed as a linear map from $\mathbb{R}^J$ to $\mathbb{R}^I$, defined by $(Mf)(i) = \sum_{j \in J} M_{i,j} f_j = M_{i,.} f$, $i \in I$, for each $f \in \mathbb{R}^J$. If $S \in \mathbb{R}^{J \times K}$ and K is finite, then the product $MS \in \mathbb{R}^{I \times K}$ corresponds to the linear map $\mathbb{R}^K \to \mathbb{R}^I$ obtained by composing the linear map corresponding to S with the linear map corresponding to M.

4.0.3 Notation from the theory of formal languages

The notation described below is standard in formal languages and automata theory, see [14,10]. Consider a finite set X which will be called the *alphabet*. Denote by X^* the set of finite sequences of elements of X. Finite sequences of elements of X will be referred to as *strings* or *words* over the alphabet X. For a word $w = a_1 a_2 \cdots a_k \in X^*$, $a_1, a_2, \ldots, a_k \in X$, $k > 0$ the length of w is denoted by $|w|$, i.e. $|w| = k$. We will denote by ϵ the *empty sequence (word)*. The length of the empty sequence ϵ is zero: $|\epsilon| = 0$. We will denote by X^+ the set of of non-empty words over X. That is, $X^+ = X^* \setminus \{\epsilon\}$. Consider two words $w \in X^*$ and $v \in X^*$ of the form $v = v_1 v_2 \cdots v_k$, and $w = w_1 w_2 \cdots w_m$, $v_1, v_2, \ldots, v_k, w_1, w_2, \ldots, w_m \in X$. Define the concatenation $vw \in X^*$ of the words v and w as the the word $vw = v_1 v_2 \cdots v_k w_1 w_2 \cdots w_m$. In particular, if $v = \epsilon$, i.e. if $k = 0$, then $vw = w$. Similarly, if $w = \epsilon$, i.e. $m = 0$, then $vw = v$. If $w \in X^+$ is a word, then w^k denotes the word $\underbrace{ww \cdots w}_{k-times}$. The word w^0 is just the empty word ϵ.

4.1 Definition and basic properties of linear switched systems

Below we present the formal definition of LSSs. For a more detailed exposition, see [51,35,45,44,42].

Definition 1 (Linear switched systems) *A* linear switched system *(ab-*

breviated by LSS) is a control system Σ of the form

$$\Sigma : \begin{cases} \dot{x}(t) = & A_{q(t)}x(t) + B_{q(t)}u(t) \\ y(t) = & C_{q(t)}x(t) \end{cases} \tag{5}$$

- $x(t) \in \mathcal{X}$ *is the continuous state at time* $t \in T$*,*
- $u(t) \in \mathcal{U}$ *denotes the continuous input at time* $t \in T$*,*
- $q(t) \in Q$ *denotes the discrete mode (state) at time* t*,*
- $y(t) \in \mathbb{R}^p$ *denotes the continuous output at time* $t \in T$*.*
- *The state-space is* $\mathcal{X} = \mathbb{R}^n$*, the input-space is* $\mathcal{U} = \mathbb{R}^m$*, the output-space is* $\mathcal{Y} = \mathbb{R}^p$*, and* Q *is the finite set of discrete modes (discrete states). Here* n, m, p *are positive integers.*
- *For each discrete mode* $q \in Q$*, the corresponding matrices are of the form* $A_q \in \mathbb{R}^{n \times n}$*,* $B_q \in \mathbb{R}^{n \times m}$ *and* $C_q \in \mathbb{R}^{p \times n}$*.*

We will use $(\mathcal{X}, \mathcal{U}, \mathcal{Y}, Q, \{(A_q, B_q, C_q) \mid q \in Q\})$ *as a short-hand notation for LSSs of the form (5).*

Notation 1 (Notation for the spaces of inputs and outputs) *In the sequel we denote by* $\mathcal{U}$ *the space* $\mathbb{R}^m$ *of continuous-valued input, by* $\mathcal{Y}$ *the space* $\mathbb{R}^p$ *of continuous-valued outputs, and by* Q *the set of discrete modes.*

Informally, the system (5) evolves as follows. For a piecewise-constant *switching signal* $q(.) : T \to Q$ and piecewise-continuous input $u : T \to \mathcal{U}$, the *state trajectory* $x : T \to \mathcal{U}$ is a continuous piecewise-differentiable function which satisfies the differential equation (5). The output at time $t \in T$ is obtained by applying to $x(t)$ the readout map (matrix) $C_{q(t)}$. Below we define state- and output- trajectories more rigorously. To this end, we define the notion of switching sequences.

Definition 2 (Switching sequences) *A* switching sequence *is a sequence of the form* $w = (q_1, t_1)(q_2, t_2) \cdots (q_k, t_k)$*, where* $q_1, q_2, \ldots, q_k \in Q$ *are discrete modes and* $t_1, t_2, \ldots, t_k$ *denote the* switching times *and* $k \geq 0$*. The set of all switching sequences is denoted by* $(Q \times T)^*$*. If* $k = 0$ *then we say that* w *is the empty switching sequence and we denote it by* ϵ*. We denote the set of all* non-empty switching sequences *by* $(Q \times T)^+$*.*

The interpretation of the sequence $w = (q_1, t_1)(q_2, t_2) \cdots (q_k, t_k)$ is the following. From time instance 0 to time instance t_1 the active discrete mode is q_1, i.e. the value of the switching signal is q_1, from t_1 to $t_1 + t_2$ the value of the switching signal is q_2, from $t_1 + t_2$ to $t_1 + t_2 + t_3$ the value of the switching signal is q_3, and so on. That is, the non-negative real t_i indicates the time spent in the discrete mode q_i, for all $i = 1, 2, \ldots, k$. *In this paper the switching sequences are regarded as inputs and we allow any switching sequence to occur.*

Definition 3 (State of LSSs) *Let $u \in PC(T,\mathcal{U})$ be a continuous-valued input and let $w = (q_1,t_1)(q_2,t_2)\cdots(q_k,t_k) \in (Q \times T)^+$ be a non-empty switching sequence. The* state of Σ reached from the initial state $x_0 \in \mathcal{X}$ with the inputs u and w *is denoted by $x_\Sigma(x_0,u,w)$ and it is defined recursively on k as follows.*

- *If $k = 1$, then $x_\Sigma(x_0,u,(q_1,t_1))$ is the solution at time t_1 of the differential equation*
$$\dot{x}(s) = A_{q_1}x(s) + B_{q_1}u(s)$$
with the initial condition $x(0) = x_0$.
- *If $x_\Sigma(x_0,u,(q_1,t_1)(q_2,t_2)\cdots(q_{k-1},t_{k-1}))$ is already defined, then let $x_\Sigma(x_0,u,w)$ be the solution at time t_k of the differential equation*
$$\dot{x}(s) = A_{q_k}x(s) + B_{q_k}u(s + \sum_{j=1}^{k-1} t_j)$$
with the initial condition $x(0) = x_\Sigma(x_0,u,(q_1,t_1)\cdots(q_{k-1},t_{k-1}))$.

That is, the states evolves according to the differential equation determined by the discrete mode. If a switch occurs, then the state at the time of the switch is taken as the initial condition for generating a solution to the differential equation associated with the new discrete mode. As a next step, we will define the output trajectories of LSSs.

Definition 4 (Output of LSSs) *Consider a continuous-valued input $u \in PC(T,\mathcal{U})$ and a non-empty switching sequence $w = (q_1,t_1)(q_2,t_2)\cdots(q_k,t_k) \in (Q \times T)^+$. The* output generated by the LSS Σ if it is started from initial state $x_0 \in \mathcal{X}$ and fed with the inputs u and w, *denoted by $y_\Sigma(x_0,u,w) \in \mathcal{Y}$, is defined by*

$$y_\Sigma(x_0,u,w) = C_{q_k}x_\Sigma(x_0,u,w) \tag{6}$$

That is, the current output is obtained from the current continuous state $x_\Sigma(x_0,u,w)$ by the application of the readout map C_{q_k} associated with the current discrete mode q_k. We define the input-output map of a LSS induced by a particular initial state as follows.

Definition 5 (Input-output maps of LSSs) *Consider a state $x_0 \in \mathcal{X}$ of Σ. Define the* input-output map of Σ induced by the state x_0 *as the map $y_\Sigma(x_0,.,.) : PC(T,\mathcal{U}) \times (Q \times T)^+ \to \mathcal{Y}$ such that for all continuous-valued input $u \in PC(T,\mathcal{U})$ and for all non-empty switching sequence $w \in (Q \times T)^+$*

$$y_\Sigma(x_0,.,.)(u,w) = y_\Sigma(x_0,u,w) \tag{7}$$

That is, the value of $y_\Sigma(x_0,.,.)$ at (u,w) equals the output generated by Σ if started from the initial states x_0 and fed inputs u and w. By abuse of notation we will denote $y_\Sigma(x_0,.,.)(u,w)$ by $y_\Sigma(x_0,u,w)$.

5 Overview of realization theory for LSS

The goal of this section is to present a brief overview of realization theory of LSSs. In Subsection 5.1 we recall the definition of input-output maps of LSSs and the notion of a LSS realization. In Subsection 5.2. we review a number of system-theoretic concepts such as observability, semi-reachability, minimality and LSS morphisms. In Subsection 5.3 the realization problem will be formulated. Subsection 5.4 recalls the concept of generalized kernel representation of a family of input-output maps. Subsection 5.5 presents the definition of Markov-parameters for LSSs. Finally, Subsection 5.6 presents the main results on realization theory of LSSs. For a more details on the material of this section see [44,42,45].

5.1 Input-output maps

In this section we define the class of maps which represent potential input-output maps of LSSs. In addition, we introduce the notion of a LSS realization.

Definition 6 (Input-output maps) *In this paper, unless stated otherwise, an* input-output map *will mean a map of the form* $f : PC(T,\mathcal{U}) \times (Q \times T)^+ \rightarrow \mathcal{Y}$*. The set of all such maps will be denoted by* $F(PC(T,\mathcal{U}) \times (Q \times T)^+, \mathcal{Y})$*. A* family of input-output maps *is just a (possibly infinite) subset of the set of all input-output maps* $F(PC(T,\mathcal{U}) \times (Q \times T)^+, \mathcal{Y})$*.*

That is, an input-output map maps continuous-valued inputs and non-empty switching sequences to continuous-valued outputs. In order to formalize the notion of a realization by a LSS of a family of input-output maps we will adopt the following formalism.

Definition 7 (Realization of input-output maps) *Consider a set* $\Phi \subseteq F(PC(T,\mathcal{U}) \times (Q \times T)^+, \mathcal{Y})$ *of input-output maps. The family* Φ *is said to be* realized *by a LSS* Σ *of the form (5) if there exists a map* $\mu : \Phi \rightarrow \mathcal{X}$*, which maps each input-output map* f *from* Φ *to a state* $\mu(f)$ *of* Σ*, such that* f *equals the input-output map induced by* $\mu(f)$*, i.e.*

$$\forall u \in PC(T,\mathcal{U}), w \in (Q \times T)^+ : y_{\Sigma}(\mu(f), u, w) = f(u, w) \tag{8}$$

One can think of the map μ as a way to determine the initial state corresponding to each element of Φ.

Definition 8 (Realizations) *Let* Φ *be a family of input-output maps. We will refer to the pairs* (Σ, μ)*, where* Σ *is a LSS of the form (5) and* $\mu : \Phi \rightarrow$

$\mathcal{X}$ is a map mapping elements of Φ to the states of Σ, as LSS realizations *(*realizations *for short). The realization (Σ, μ) is said to be a* realization of Φ*, if (8) holds for all $f \in \Phi$.*

Note that not any realization (Σ, μ) with $\mu : \Phi \to \mathcal{X}$ is a realization of Φ. The statement that (Σ, μ) is a *realization* only expresses the fact that we associate a state of Σ with each element of Φ. However, we do not yet require that the input-output map induced by a designated state equals the corresponding element of Φ. The latter is required only if we claim that (Σ, μ) *is a realization of* Φ.

5.2 System-theoretic concepts

The goal of this section is to define system theoretic concepts such as observability, span-reachability, system morphism, dimension and minimality for LSSs and for LSS realizations. Throughout the section Σ denotes a LSS of the form (5).

The *reachable set* of Σ from a set of initial states $\mathcal{X}_0 \subseteq \mathcal{X}$ is defined as

$$Reach(\Sigma, \mathcal{X}_0) = \{x_\Sigma(x_0, u, w) \in \mathcal{X} \mid u \in PC(T, \mathcal{U}), w \in (Q \times T)^+, x_0 \in \mathcal{X}_0\}$$

That is, $Reach(\Sigma, \mathcal{X}_0)$ is the set of all those states which are obtained by starting the system from an initial state in $\mathcal{X}_0$, applying some continuous-valued input and some finite switching sequence, and considering the state at the last switching time.

Definition 9 ((Semi-)Reachability) *The LSS Σ is said to be* reachable *from $\mathcal{X}_0$ if $Reach(\Sigma, \mathcal{X}_0) = \mathcal{X}$ holds. The LSS Σ is* semi-reachable *from $\mathcal{X}_0$ if $\mathcal{X}$ is the smallest vector space containing $Reach(\Sigma, \mathcal{X}_0)$.*

In other words, Σ is semi-reachable from $\mathcal{X}_0$ if the linear span of the elements of the reachable set $Reach(\Sigma, \mathcal{X}_0)$ yields the whole state-space $\mathcal{X}$. We proceed with defining the notion of observability for LSSs.

Definition 10 (Observability and Indistinguishability) *Two states $x_1 \neq x_2 \in \mathcal{X}$ of the LSS Σ are* indistinguishable *if the input-output maps induced by x_1 and x_2 coincide, i.e. for all $u \in PC(T, \mathcal{U})$ and $w \in (Q \times T)^+$, $y_\Sigma(x_1, u, w) = y_\Sigma(x_2, u, w)$. The LSS Σ is called* observable *if it has no pair of distinct indistinguishable states.*

In other words, observability means that if we pick any two states of the system, then we are able to distinguish between them by feeding a suitable continuous-valued input and a suitable switching sequence and then observing the resulting output.

Below we will define the notion of dimension for LSSs.

Definition 11 (Dimension of LSSs) *Define the dimension of* Σ*, denoted by* $\dim \Sigma$*, as the dimension* $\dim \mathcal{X} = n$ *of is state-space.*

Now we are ready to define the concept of minimality for LSSs .

Definition 12 (Minimality of LSSs) *Let* Φ *be a family of input-output maps and let* (Σ, μ) *be a LSS realization of* Φ*.* (Σ, μ) *is* a minimal realization of Φ*, if for any LSS realization* $(\hat{\Sigma}, \hat{\mu})$ *of* Φ*,* $\dim \Sigma \leq \dim \hat{\Sigma}$*.*

In simple words, a LSS realization is a minimal realization of Φ if it has the smallest dimensional state-space among all the linear switched systems which are realizations of Φ.

The notions of observability and semi-reachability can be extended to LSS realizations as follows.

Definition 13 (Observability and semi-reachability of realizations) *Let* Φ *be a family of input-output maps and let* $\mu : \Phi \to \mathcal{X}$ *be a map from* Φ *to the state-space of* Σ*. The realization* (Σ, μ) *is* semi-reachable *, if* Σ *is semi-reachable from the range* $\mathrm{Im}\mu$ *of* μ*. The realization* (Σ, μ) *is* observable*, if* Σ *is observable.*

As the next step, we will define the notion of a LSS morphism.

Definition 14 (Linear switched system morphism) *Consider a LSS* Σ_1 *of the form (5) and a LSS* Σ_2 *of the form* $\Sigma_2 = (\mathcal{X}_a, \mathcal{U}, \mathcal{Y}, Q, \{(A^a_q, B^a_q, C^a_q) \mid q \in Q\})$ [1] *. A linear map* $S : \mathcal{X} \to \mathcal{X}_a$ *is said to be a* LSS morphism *from* Σ_1 *to* Σ_2*, and it is denoted by* $S : \Sigma_1 \to \Sigma_2$*, if for all discrete modes* $q \in Q$*,*

$$A^a_q S = S A_q,\ B^a_q = S B_q,\ C^a_q S = C_q \tag{9}$$

The map S *is called surjective (injective) if it is surjective (injective) as a linear map. The map* S *is said to be a LSS isomorphism, if it is an isomorphism as a linear map. Consider two LSS realizations* (Σ_1, μ_1) *and* (Σ_2, μ_2) *such that the domain of definition of both* μ_1 *and* μ_2 *is a certain family* Φ *of input-output maps. A LSS morphism* $S : \Sigma_1 \to \Sigma_2$ *is called a* LSS morphism *from realization* (Σ_1, μ_1) *to* (Σ_2, μ_2)*, if* $S \circ \mu_1 = \mu_2$ *holds, or, in other words, if for all* $f \in \Phi$*,* $S(\mu_1(f)) = \mu_2(f)$*. The fact that* S *is a LSS morphism from* (Σ_1, μ_1) *to* (Σ_2, μ_2) *will be denoted by* $S : (\Sigma_1, \mu_1) \to (\Sigma_2, \mu_2)$*. The LSSs realizations* (Σ_1, μ_1) *and* (Σ_2, μ_2) *are said to be* algebraically similar *or* isomorphic *if there exists an LSS isomorphism* $S : (\Sigma_1, \mu_1) \to (\Sigma_2, \mu_2)$*.*

[1] Notice that the LSSs Σ_1 and Σ_2 have the same set of discrete modes.

5.3 Realization problem

The realization problem for LSSs can be formulated as follows.

Problem 1 (Realization problem for LSSs) *Find necessary and sufficient conditions for existence of a LSS realization for a family of input-output maps Φ. Find a characterization of minimal LSS realizations of Φ. Determine if minimal realizations of Φ are unique in any sense.*

5.4 Generalized kernel representation

In this section we will recall the notion of *generalized kernel representation*. It turns out that a family of input-output maps can be realized by a LSS only if it admits a generalized kernel representation. Informally, a family Φ of input-output maps has a generalized kernel representation if the following hold.

(1) There exists an input-output map y^{Φ} such that for all $f \in \Phi$, $f(u,w) = f(0,w) + y^{\Phi}(u,w)$ for all continuous-valued inputs u and switching sequences w.
(2) Each element f of Φ is affine in continuous-valued inputs and analytic in switching times for all constant inputs.

A good intuition for the notion of generalized kernel representation can be derived by analogy with input-output maps of linear systems. Recall from [6] that an input-output map $y : PC(T,\mathcal{U}) \times T \to \mathcal{Y}$ can be realized by a linear system $(A,B,C) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times m} \times \mathbb{R}^{p \times n}$ from the initial state $x_0 \in \mathbb{R}^n$, only if there exist analytic functions $K : T \to \mathbb{R}^p$ and $G : T \to \mathbb{R}^{p \times m}$ such that

$$y(u,t) = K(t) + \int_0^t G(t-s)u(s)ds \tag{10}$$

More precisely, in this case

$$K(t) = Ce^{At}x_0 \quad \text{and} \quad G(t) = Ce^{At}B \tag{11}$$

Using the setting above, the maps $K_w^{f,\Phi}$ and G_w^{Φ} to be defined below are analogous to the map K and G respectively. The formal definition is as follows.

Definition 15 (Generalized kernel-representation) *Consider a family of input-output maps Φ, as defined in Definition 6. The family Φ is said to have* generalized kernel representation*, if for all input-output maps $f \in \Phi$ and for all non-empty sequences of discrete modes $w = q_1 q_2 \cdots q_k \in Q^+$, $q_1, q_2 \ldots, q_k \in Q$,*

$k > 0$, there exist functions

$$K^{f,\Phi}_w : T^k \rightarrow \mathbb{R}^p \quad \text{and } G^{\Phi}_w : T^k \rightarrow \mathbb{R}^{p \times m}$$

such that the following holds.

(1) For each word $w \in Q^+$ and for each input-output map $f \in \Phi$, the functions $K^{f,\Phi}_w$ and G^{Φ}_w are analytic.

(2) For each input-output map $f \in \Phi$ and for each (possibly empty) sequences $w, v \in Q^*$, and for each discrete mode $q \in Q$, it holds that for all $t_1, t_2, \ldots, t_{|w|}, t, \hat{t}, t_{|w|+2}, \ldots, t_{|w|+|v|+1} \in T$,

$$\begin{aligned}
&K^{f,\Phi}_{wqqv}(t_1, t_2, \ldots, t_{|w|}, t, \hat{t}, t_{|w|+2}, \ldots t_{|w|+|v|+1}) = \\
&= K^{f,\Phi}_{wqv}(t_1, t_2, \ldots t_{|w|}, t + \hat{t}, t_{|w|+2} \ldots t_{|w|+|v|+1}) \\
&G^{\Phi}_{wqqv}(t_1, t_2, \ldots, t_{|w|}, t, \hat{t}, t_{|w|+2}, \ldots t_{|w|+|v|+1}) = \\
&= G^{\Phi}_{wqv}(t_1, t_2, \ldots t_{|w|}, t + \hat{t}, t_{|w|+2} \ldots t_{|w|+|v|+1})
\end{aligned}$$

(3) For each pair of sequences $v, w \in Q^*$ such that and w is not the empty word, i.e. $|w| > 0$, and for each discrete mode $q \in Q$, the following holds. For each input-output map $f \in \Phi$ and for each $t_1, t_2, \ldots, t_{|vw|} \in T$,

$$K^{f,\Phi}_{vqw}(t_1, t_2, \ldots, t_{|v|}, 0, t_{|v|+1}, \ldots, t_{|wv|}) = K^{f,\Phi}_{vw}(t_1, t_2, \ldots, t_{|vw|})$$

For each pair of words $v, w \in Q^*$ such that both v and w are not empty, i.e. $|v| > 0$, $|w| > 0$, and for each discrete mode $q \in Q$, the following holds. For each $t_1, t_2, \ldots, t_{|vw|} \in T$,

$$G^{\Phi}_{vqw}(t_1, t_2, \ldots, t_{|v|}, 0, t_{|v|+1}, \ldots, t_{|wv|}) = G^{\Phi}_{vw}(t_1, t_2, \ldots, t_{|vw|})$$

(4) For each input-output map $f \in \Phi$, for each non-empty switching sequence $w = (q_1, t_1)(q_2, t_2) \cdots (q_k, t_k)(Q \times T)^+$, where $q_1, q_2, \ldots, q_k \in Q$ and $t_1, t_2, \ldots, t_k \in T$, each piecewise-continuous input $u \in PC(T, \mathcal{U})$, the following holds.

$$f(u, w) = K^{f,\Phi}_{q_1 q_2 \cdots q_k}(t_1, t_2, \ldots, t_k) + \\
\sum_{i=1}^{k} \int_0^{t_i} G^{\Phi}_{q_i q_{i+1} \cdots q_k}(t_i - s, t_{i+1}, \ldots, t_k) u(s + \sum_{j=1}^{i-1} t_j) ds$$

The reader may view the functions $K^{f,\Phi}_w$ as the part of the output which depends on the initial condition and the functions G^{Φ}_w as functions determining the dependence of the output on the continuous inputs.

In fact, let (Σ, μ) be a LSS realization of Φ and assume that Σ is of the form (5). It is easy to see that for all $f \in \Phi$ and for any input $u \in PC(T, \mathcal{U})$ and switching sequence $(q_1,t_1)(q_2,t_2)\cdots(q_k,t_k) \in (Q\times T)^+$,

$$\begin{aligned}
&y_{\Sigma}(\mu(f),u,w) = C_{q_k}e^{A_{q_k}t_k}e^{A_{q_{k-1}}t_{k-1}}\cdots e^{A_{q_1}t_1}\mu(f)+\\
&+\int_0^{t_k} C_{q_k}e^{A_{q_k}(t_k-s)}B_{q_k}u(s+\sum_1^{k-1}t_i)ds+\\
&+C_{q_k}e^{A_{q_k}t_k}\int_0^{t_{k-1}} e^{A_{q_{k-1}}(t_{k-1}-s)}B_{q_{k-1}}u(s+\sum_1^{k-2}t_i)ds+\cdots\\
&\cdots+C_{q_k}e^{A_{q_k}t_k}e^{A_{q_{k-1}}t_{k-1}}\cdots e^{A_{q_2}t_2}\int_0^{t_1} e^{A_{q_1}(t_1-s)}B_{q_1}u(s)ds
\end{aligned} \tag{12}$$

From the equation above it is easy to see that Φ admits a hybrid kernel representation of the form

$$\begin{aligned}
G^{\Phi}_{q_1q_2\cdots q_k}(t_1,t_2,\dots,t_k) =&C_{q_k}e^{A_{q_k}t_k}e^{A_{q_{k-1}}t_{k-1}}\cdots e^{A_{q_1}t_1}B_{q_1}\\
K^{f,\Phi}_{q_1q_2\cdots q_k}(t_1,t_2,\dots,t_k) =&C_{q_k}e^{A_{q_k}t_k}e^{A_{q_{k-1}}t_{k-1}}\cdots e^{A_{q_1}t_1}\mu(f).
\end{aligned} \tag{13}$$

for all non-empty sequences of discrete modes $q_1,q_2,\dots,q_k \in Q$, $k\ge 1$.

5.5 Generalized Markov parameters for LSS

Next we will define the notion of Markov parameters for input-output maps. Markov-parameters play a central role in (partial) realization theory of LSSs.

Before proceeding further, recall from classical linear systems theory [6] the notion of Markov parameter. Consider the linear input-output map of the form (10) and define the Markov parameters of this map as derivatives of the maps $K: T\to\mathbb{R}^p$ and $G: T\to\mathbb{R}^{p\times m}$; $M_k = \frac{d^k}{dt^k}G(t)|_{t=0}$ and $N_k = \frac{d^k}{dt^k}K(t)|_{t=0}$ for all $k\ge 0$. In turn, the derivatives of $G(t)$ and $K(t)$ can be expressed as the derivatives of the input-output map as follows. For a constant input $u\in\mathcal{U}$ define the map

$$f_u : T \ni t \mapsto K(t) + (\int_0^t G(t-s)ds)u$$

That is, $f_u(t)$ is just the value of the input-output map (10) at time t if a constant input u is fed in. Then $f_0(t) = K(t)$ and hence $N_k = \frac{d^k}{dt^k}f_0(t)|_{t=0}$. Similarly, the jth column of M_k can be written as $\frac{d^{k+1}}{dt^{k+1}}(f_{e_j}(t)-f_0(t))|_{t=0}$ where e_j, $j=1,2,\dots,m$ is the jth unit vector of $\mathcal{U}=\mathbb{R}^m$. That is, *the Markov-parameters of a linear input-output map are the high-order time derivatives of the output trajectories induced by certain constant inputs, evaluated at zero.*

For LSSs the Markov-parameters are defined in a similar way. Before proceeding to the definition, we need the following notation

Notation 2 (Input-output maps as time functions) *Consider an input-output map f as in Definition 6, a non-empty sequence of discrete modes*

$w = q_1 q_2 \cdots q_k \in Q^+$, $q_1, q_2, \ldots, q_k \in Q$, $k \geq 1$ *and a continuous-valued input* $u \in PC(T,\mathcal{U})$. *Define the map* $f_{u,w} : T^k \to \mathcal{Y}$ *as follows*

$$f_{u,w}(t_1, t_2, \ldots, t_k) = f(u, (q_1, t_1)(q_2, t_2), \ldots, (q_k, t_k)) \tag{14}$$

That is, the values of $f_{u,w}$ are obtained from the values of f by fixing the a piecewise-continuous input u and a sequence of discrete modes w and varying the switching times only.

Definition 16 (Markov-parameters of Φ) *Let Φ be a family of input-output maps admitting a generalized kernel representation. The* Markov parameters *of Φ are the vectors $S_{q_0,q,j}(w) \in \mathbb{R}^p$ and $S_{f,q}(w) \in \mathbb{R}^p$, defined for all discrete modes $q_0, q \in Q$, all the sequences of discrete modes $w \in Q^*$ (including the empty sequence), all $j = 1, 2, \ldots, m$ and $f \in \Phi$ as follows. If $w = q_1 q_2 \cdots q_k$ for $k \geq 0$, $q_1, q_2, \ldots, q_k \in Q$, then*

$$\begin{aligned}
S_{f,q}(w) =& \frac{d}{dt_1}\frac{d}{dt_2}\cdots\frac{d}{dt_k} f_{0,q_1q_2\cdots q_k q}(t_1, t_2, \ldots, t_k, 0)|_{t_1=t_2\cdots t_k=0} \\
S_{q,q_0,j}(w) =& \frac{d}{dt_0}\frac{d}{dt_1}\frac{d}{dt_2}\cdots\frac{d}{dt_k} f_{e_j,q_0q_1q_2\cdots q_k q}(t_0, t_1, t_2, \ldots, t_k, 0)|_{t_0=t_1=t_2\cdots t_k=0} \\
&- \frac{d}{dt_0}\frac{d}{dt_1}\frac{d}{dt_2}\cdots\frac{d}{dt_k} f_{0,q_0q_1q_2\cdots q_k q}(t_0, t_1, t_2, \ldots, t_k, 0)|_{t_0=t_1=t_2\cdots t_k=0}
\end{aligned}$$

Here e_j is the jth unit vector of $\mathcal{U} = \mathbb{R}^m$.

The vectors $S_{f,q}(w)$ correspond to the parameters N_k for linear systems, and $S_{q,q_0,j}(w)$ corresponds to the jth column of the M_k component of the Markov parameters. As it was shown in [44,42,45], there is a close relationship between the Markov parameters of Φ and products of system matrices.

Lemma 1 ([44,45,42]) *Let Σ be of the form (5) and let $\mu : \Phi \to \mathcal{X}$. The realization (Σ, μ) is a realization of Φ, if and only if the Markov-parameters of Φ are equal to the following products of the system matrices.*

$$\begin{aligned}
S_{q,q_0,j}(q_1 q_2 \cdots q_k) &= C_q A_{q_k} A_{q_{k-1}} \cdots A_{q_1} B_{q_0} e_j \\
S_{f,q}(q_1 q_2 \cdots q_k) &= C_q A_{q_k} A_{q_{k-1}} \cdots A_{q_1} \mu(f)
\end{aligned} \tag{15}$$

for each $q, q_0 \in Q$, for each $q_1, q_2, \ldots, q_k \in Q$, $k \geq 0$, $f \in \Phi$ and $j = 1, 2, \ldots, m$.

Similarly to the linear case, the Markov-parameters of Φ can also be represented as derivatives of the maps $K_w^{f,\Phi}$ and G_w^{Φ}, see [44,42,45].

The purpose of this section is to present formally the main results on realization theory of LSSs . To this end, the notion of the Hankel-matrix H_Φ of Φ is defined. Similarly to the linear case, the entries of the Hankel-matrix will be formed by the Markov parameters.

Definition 17 (Hankel-matrix) *Assume that the set of discrete modes Q has D elements and choose an enumeration of Q*

$$Q = \{\sigma_1, \sigma_2, \dots, \sigma_D\} \tag{16}$$

Recall the notation for infinite matrices presented in Section 4.0.2. Define the Hankel-matrix *of Φ as the infinite real matrix, columns and rows of which are indexed as follows. The rows of H_Φ are indexed by pairs (v, i) where $v \in Q^*$ is a finite sequence of discrete modes and $i \in \{1, 2, \dots, pD\}$. The columns of H_Φ are indexed by pairs (w, j), where $w \in Q^*$ is a sequence of discrete modes and j is either an element of Φ, or j is a pair of the form (q, z), where q is a discrete mode and $z \in \{1, 2, \dots, m\}$. That is, H_Φ is an infinite real matrix of the form*

$$H_\Phi \in \mathbb{R}^{(Q^* \times I) \times (Q^* \times J_\Phi)} \tag{17}$$

where $I = \{1, 2, \dots, pD\}$ and $J_\Phi = \Phi \cup (Q \times \{1, 2, \dots, m\})$. The entries of H_Φ are defined as follows. Fix sequences of discrete modes $w, v \in Q^*$ and fix an element $j \in J_\Phi$. For any i of the form $i = pK + r + 1$ where $K = 0, 1, \dots, D-1$ and $r = 0, 1, \dots, p-1$, the entry $(H_\Phi)_{(v,i),(w,j)}$ is defined as follows

$$(H_\Phi)_{(v,i),(w,j)} = \begin{cases} (S_{\sigma_{K+1},q,z}(wv))_{r+1} & \text{if } j = (q, z) \in Q \times \{1, \dots, m\} \\ (S_{f,\sigma_{K+1}}(wv))_{r+1} & \text{if } j = f \in \Phi \end{cases} \tag{18}$$

Here $(S_{\sigma_{K+1},q,z}(wv))_{r+1}$ and $(S_{f,\sigma_{K+1}}(wv))_{r+1}$ denote the $r+1$th element of the vectors $S_{\sigma_{K+1},q,z}(wv) \in \mathbb{R}^p$ and $S_{f,\sigma_{K+1}}(wv) \in \mathbb{R}^p$ respectively.

That is, H_Φ is constructed from certain high-order derivatives of the input-output maps belonging to Φ. As it was noted in Section 4.0.2, the columns of H_Φ belong to the vector space of all maps $(Q^* \times \{1, 2, \dots, pD\}) \to \mathbb{R}$. Hence, we can speak of the linear span of the columns of H_Φ. In addition, according to the convention adopted in Section 4.0.2, the *rank of* H_Φ, denoted by rank $H_\Phi \in \mathbb{N} \cup \{+\infty\}$, is the dimension of the linear subspace spanned by the columns of H_Φ. Now we are ready to state the main theorem on the existence of a LSS realization for arbitrary switching.

Theorem 1 (Realization of input-output maps, [44,45,42]) *Let Φ be a family of input-output maps. Then Φ has a realization by a LSS if and only* *if* Φ *has a generalized kernel representation and the rank of the associated Hankel-matrix* H_Φ *of* Φ *is finite, i.e,* $\text{rank } H_\Phi < +\infty$.

Next, we state state the main result of on minimality of LSSs .

Theorem 2 (Minimality, [44,45,42]) *If* (Σ, μ) *is a LSS realization of* Φ, *then the following are equivalent.*

(i) (Σ, μ) *is a minimal LSS realization of* Φ.
(ii) The realization (Σ, μ) *is semi-reachable and it is observable.*
(iii) The state-space dimension of Σ *equals the rank of the Hankel-matrix of* Φ, *i.e.* $\dim \Sigma = \dim H_\Phi$.

In addition, all minimal LSS realizations of Φ *are isomorphic.*

6 Main results of the paper

The purpose of this section is to present the main results of the paper formally. The outline of this section is as follows. In Subsection 6.1 we state the partial-realization problem for LSSs formally. In Subsection 6.2 we define the notion of Hankel sub-matrix, which will be needed for the statement of the main results. In Subsection 6.3 we present the theorem characterizing the existence and minimality of partial realizations. In addition, we present a simple algorithm for computing a minimal partial realization. Finally, in Subsection 6.4 we present a Kalman-Ho-like algorithm for computing a minimal partial realization.

6.1 Partial-realization problem

Let Φ be a family of input-output maps admitting a generalized kernel representation. As we saw earlier, the Markov-parameters uniquely determine the input-output maps of a linear switched system. In addition, the realization problem for LSS can be reduced to the problem of finding a suitable representation of the Markov-parameters of the input-output maps. However, in practice we can obtain only Markov-parameters only up to some finite order. Hence, the *partial-realization problem* (with respect to the number of switches) arises. In order to present a formal problem formulation, the notion of (minimal) partial realization has to be introduced.

Definition 18 (N-partial realization) *Let* Φ *be a family of input-output maps admitting a generalized kernel representation. Assume that* Σ *is a LSS* *of the form (5) and let $\mu : \Phi \to \mathcal{X}$. A realization (Σ, μ) is said to be an N-partial realization of Φ, if for all $q, q_0 \in Q$, $f \in \Phi$, $j = 1, 2, \dots, m$, and for all (possibly empty) sequences of discrete modes $q_1, q_2, \dots, q_k \in Q$, of length at most N, i.e. $N \geq k \geq 0$, the following holds.*

$$\begin{aligned} S_{q,q_0,j}(q_1 q_2 \cdots q_k) =& C_q A_{q_k} A_{q_{k-1}} \cdots A_{q_1} B_{q_0} e_j \\ S_{f,q}(q_1 q_2 \cdots q_k) =& C_q A_{q_k} A_{q_{k-1}} \cdots A_{q_1} \mu(f) \end{aligned} \tag{19}$$

The LSS Σ is said to be an N-partial realization of Φ, if for there exists a map $\mu : \Phi \to \mathcal{X}$ such that the realization (Σ, μ) is a N-partial realization of Φ. The realization (Σ, μ) is said to be a partial realization of Φ, if it is a N-partial realization of Φ for some $N \in \mathbb{N}$.

From Lemma 1 it follows that (Σ, μ) is a realization of Φ if and only if (Σ, μ) is an N-partial realization of Φ for all $N \in \mathbb{N}$. Note that if Φ and (Σ, μ) is a N-partial realization of Φ, then (19) holds for *finitely many* Markov-parameters.

Remark 1 (Terminology: Markov-parameters of order N) *In the sequel, we will often refer to the Markov-parameters which are indexed by a sequence of discrete modes of length N as* Markov-parameters of order N. *That is, a Markov-parameter of order N is either of the form $S_{q,q_0,j}(w)$ or $S_{f,q}(w)$ with $|w| = N$.*

With this terminology, a LSSis a N-partial realization of Φ, if it recreates the Markov-parameters of Φ of order at most N. If Φ is finite, then the set of Markov-parameters of Φ of order at most N is finite.

Definition 19 (Minimal N-partial realization) *A LSS realization (Σ, μ) is said to be a minimal N-partial realization of Φ, if (Σ, μ) is a N-partial realization of Φ and for any other N-partial LSS realization $(\Sigma^{'}, \mu^{'})$ of Φ, $\dim \Sigma \leq \dim \Sigma^{'}$.*

In plain words, a minimal N-partial realization of Φ is a N-partial realization of Φ with the smallest possible state-space dimension.

Problem 2 (Partial-realization problem) *Let Φ be a family of input-output maps admitting a generalized kernel representation. The partial-realization problem entails the following problems.*

- *Find conditions for existence of a N-partial realization of Φ. Formulate an algorithm for computing a N-partial realization from generalized Markov-parameters of Φ of some bounded order. If Φ is finite, this amounts to computing a N-partial realization from* finitely many *Markov-parameters.*
- *Characterize minimal N-partial realizations of Φ, find conditions for their existence and uniqueness.*

- *Find conditions under which a N-partial realization of Φ is a complete realization of Φ in the sense of Definition 8.*

We will devote the remaining part of the section to presenting the solution to the problem formulated above. We will show that it is possible to compute a minimal N-partial realization from a suitably chosen sub-matrix of the Hankel matrix of Φ. In fact, this sub-matrix is formed by Markov-parameters of bounded order, and hence it is finite, if Φ is finite. If N is large enough, the resulting N-partial realization will be a complete realization of Φ. In addition, if Φ is finite, then this sub-matrix will be finite as well.

6.2 Finite Hankel-matrices

In order to state the main results rigorously, we need to define formally. certain finite sub-matrices of the Hankel-matrix. To this end, recall from Definition 17 the definition the Hankel-matrix of Φ and of the sets $I = \{1, 2, \ldots, pD\}$, $D = |Q|$ and $J_\Phi = \Phi \cup (Q \times \{1, 2, \ldots, m\})$. Fix natural numbers $L, M \in \mathbb{N}$. We will use the following notation for sequences of discrete modes of length at most L.

Notation 3 *We will denote by $Q^{\leq L}$ the set of all sequences of discrete modes of length at most L, i.e. $Q^{\leq L} = \{w \in Q^* \mid |w| \leq L\}$.*

Notice that $Q^{\leq L}$ is a finite set of cardinality $(|Q|^{L+1}-1)/(|Q|-1)$. We will define the sub-matrix of the Hankel-matrix H_Φ indexed by sequences of bounded length as follows.

Definition 20 ($H_{\Phi,L,M}$ sub-matrices of the Hankel-matrix) *Let $H_{\Phi,L,M}$ be the sub-matrix of H_Φ formed by the intersections of all the rows indexed by an index of the form (v, i) with $i \in I$ and $v \in Q^{\leq L}$ and all the columns indexed by an index of the form (w, j) with $j \in J_\Phi$ and $w \in Q^{\leq M}$. That is, $H_{\Phi,L,M} \in \mathbb{R}^{(Q^{\leq L} \times I) \times (Q^{\leq M} \times J_\Phi)}$ and for each $(v, i) \in I \times Q^{\leq L}$ and $(w, j) \in (Q^{\leq M} \times J_\Phi)$,*

$$(H_{\Phi,L,M})_{(v,i),(w,j)} = (H_\Phi)_{(v,i),(w,j)}$$

Notice that the entries of$H_{\Phi,L,M}$ are made up of Markov-parameters of order at most $K + L$. Hence, if Φ is finite, then the set J_Φ is finite and the matrix $H_{\Phi,L,M}$ is a finite matrix. Since the entries of the Hankel-matrix were defined in terms of the generalized Markov-parameters of Φ, the same is true for the entries of $H_{\Phi,N,M}$. Writing out the definition of the entries of $H_{\Phi,L,M}$ and using the enumeration of Q fixed in (16) of Definition 17, we get that for any i of the form $i = Kp + r + 1$ where $K = 0, 1, \ldots, D - 1$ and $r = 0, 1, \ldots, p - 1$, the entry $(H_{\Phi,L,M})_{(v,i),(w,j)}$ is defined as follows

$$(H_{\Phi,L,M})_{(v,i),(w,j)} = \begin{cases} (S_{\sigma_{K+1},q,z}(wv))_{r+1} & \text{if } j = (q,z) \in Q \times \{1,\ldots,m\} \\ (S_{f,\sigma_{K+1}}(wv))_{r+1} & \text{if } j = f \in \Phi \end{cases} \tag{20}$$

Here $(S_{\sigma_{K+1},q,z}(wv))_{r+1}$ and $(S_{f,\sigma_{K+1}}(wv))_{r+1}$ denote the $r+1$th element of the vectors $S_{\sigma_{K+1},q,z}(wv) \in \mathbb{R}^p$ and $S_{f,\sigma_{K+1}}(wv) \in \mathbb{R}^p$ respectively. That is, if Φ is finite, then $H_{\Phi,L,M}$ represents a finite collection of Markov-parameters of Φ. In turn the Markov-parameters of Φ are defined via the high-order derivatives of the elements of Φ with respect to the switching times. Hence, we get that the entries of $H_{\Phi,L,M}$ are just high-order derivatives with respect to the switching times of the elements of Φ, and the order of these derivatives is bounded by $L+M$.

6.3 Solution of the partial-realization problem

As we remarked in the introduction, there are essentially two ways to construct a partial LSS realization. In Algorithm 1 we present the first one which is conceptually the simplest one. Informally, the realization (Σ_N, μ_N) returned by Algorithm 1 is defined on an isomorphic copy of the column space of $H_{\Phi,N,N+1}$. The jth column of the matrix B_q is the isomorphic copy of the column of $H_{\Phi,N,N+1}$ indexed by (q,j). The matrices C_q are such that if they are interpreted as linear maps, then each C_q maps the isomorphic copy of a column of $H_{\Phi,N,N+1}$ to its first p rows, i.e. rows indexed by $(\epsilon,1),(\epsilon,2),\ldots,(\epsilon,p)$, where ϵ denotes the empty word. The matrices A_q realize a shift on the columns of $H_{\Phi,N,N+1}$; by interpreting A_q as a map on the columns of $H_{\Phi,N,N+1}$, A_q maps the column indexed by an index $(w,j) \in Q^{\leq N} \times J_\Phi$ to the column indexed by (wq,j). Finally, the value $\mu_N(f)$ of the map μ_N for the input-output map $f \in \Phi$ equals the column of $H_{\Phi,N,N+1}$ indexed by (ϵ,f). Notice that Algorithm 1 may fail to return a realization, as (23) need not always have a solution.

Remark 2 *If Φ consists of finitely many input-output maps, then Algorithm 1 is a indeed an effective procedure and it can be implemented as a numerical algorithm.*

We state the partial-realization theorem as a theorem providing sufficient conditions for Algorithm 1 to yield a $2N+1$-partial, semi-reachable and observable realization of Φ.

Theorem 3 (Existence of partial realization of LSSs) *Let Φ be a family of input-output maps admitting a generalized kernel representation.*

- ***Existence and computability of a partial realization***

Algorithm 1

1: Assume that $n = \text{rank } H_{\Phi,N,N+1}$ and let $\mathcal{S}$ be a linear isomorphism from the column space of $H_{\Phi,N,N+1}$ to $\mathbb{R}^n$. For each column index $(w,j) \in Q^{\leq N+1} \times J_\Phi$, *denote by* $\mathbf{C}_{w,j}$ *the column of* $H_{\Phi,N,N+1}$ *indexed by* (w,j).

2: Define the LSS realization (Σ_N, μ_N) as follows.
Σ_N is a LSS of the form (5) such that

- The state-space of Σ_N is $\mathbb{R}^n$.
- For each mode $q \in Q$, the matrix $B_q \in \mathbb{R}^{n \times m}$ satisfies

$$B_q = \Big[\mathcal{S}(\mathbf{C}_{\epsilon,(q,1)}),\, \mathcal{S}(\mathbf{C}_{\epsilon,(q,2)}),\, \cdots,\, \mathcal{S}(\mathbf{C}_{\epsilon,(q,m)})\Big] \tag{21}$$

- For each mode $q \in Q$, the matrix $C_q \in \mathbb{R}^{p \times n}$ satisfies

$$C_q\mathcal{S}(\mathbf{C}_{w,j}) = \Big[(H_{\Phi,N,N+1})_{(\epsilon,1),(w,j)},\, (H_{\Phi,N,N+1})_{(\epsilon,2),(w,j)},\, \cdots,\, (H_{\Phi,N,N+1})_{(\epsilon,p),(w,j)}\Big]^T \tag{22}$$

  for each $w \in Q^{\leq N+1}$.
- For each mode $q \in Q$, the matrix $A_q \in \mathbb{R}^{n \times n}$ is the solution of the system of linear equations

$$\mathcal{S}^{-1}A_q\mathcal{S}(\mathbf{C}_{w,j}) = \mathbf{C}_{wq,j} \text{ for each } w \in Q^{\leq N}, j \in J_\Phi \tag{23}$$

  If the system of equations (23) does not have a solution, then abort.

3: For each input-output map $f \in \Phi$,

$$\mu_N(f) = \mathcal{S}(\mathbf{C}_{\epsilon,f}) \tag{24}$$

4: Return (Σ_N, μ_N), with $\Sigma_N = (\mathbb{R}^n, \mathcal{U}, \mathcal{Y}, Q, \{(A_q, B_q, C_q) \mid q \in Q\})$.

Assume that for some $\mathbb{N} \ni N > 0$*,*

$$\text{rank } H_{\Phi,N,N} = \text{rank } H_{\Phi,N+1,N} = \text{rank } H_{\Phi,N,N+1} \tag{25}$$

holds. Then Algorithm 1 returns a realization (Σ_N, μ_N)*, and* (Σ_N, μ_N) *is a semi-reachable and observable* $2N+1$*-partial realization of* Φ*.*

- ***Existence of a complete realization***
If $\text{rank } H_{\Phi,N,N} = \text{rank } H_\Phi$*, then (25) holds and the* $2N+1$*-partial realization* (Σ_N, μ_N) *is a minimal realization of* Φ*. The condition* $\text{rank } H_{\Phi,N,N} = \text{rank } H_\Phi$ *holds for a given* N*, if there exists a LSS realization* (Σ, μ) *of* Φ *such that* $\dim \Sigma \leq N+1$*.*

The proof of Theorem 3 is presented in Section 7. Theorem 3 implies that if Φ has a realization by a LSS , then the parameters of a minimal LSS realization of Φ are computable from, and hence are determined by, finitely many time derivatives with respect to the switching times.

Remark 3 (Partial versus complete realization) *The numerical example in Section 3 provides an example of a family of input-output maps Φ realizable by a LSS, such that Φ has the following property. For some N the family Φ satisfies (25) but not the condition* $\text{rank } H_{\Phi,N,N} = \text{rank } H_{\Phi}$. *That is, there is an $N \in \mathbb{N}$ such that Theorem 3 yields a partial realization of Φ which is not a complete realization. That is, the two statements of Theorem 3 indeed describe two separate cases.*

Theorem 3 allows us to formulate the following characterization of minimal partial LSS realizations.

Theorem 4 (Minimal partial realization) *With the notation of Theorem 3, if Φ satisfies (25), then the following holds.*

(1) A minimal $2N+1$ partial realization of Φ exists, in fact, the realization (Σ_N, μ_N) returned by Algorithm 1 is a minimal $2N+1$ partial realization of Φ.
(2) Any minimal $2N+1$ partial realization of Φ is semi-reachable and observable and it is of dimension $\text{rank } H_{\Phi,N,N}$.
(3) All minimal $2N+1$ partial realizations of Φ are isomorphic.

The proof of Theorem 4 is presented in Section 7.

Remark 4 *The reader might wonder why we talk about $2N+1$-partial realizations in Theorem 3 and 4. The reason behind it is that the finite Hankel-matrix $H_{\Phi,N,N+1}$ is formed by values of the Markov-parameters of order at most $2N+1$. We would like the LSS realization obtained from $H_{\Phi,N,N+1}$ to recreate at least those Markov-parameters which are the entries of the matrix $H_{\Phi,N,N+1}$. But this means that the LSS realization obtained from $H_{\Phi,N,N+1}$ must be a $2N+1$-partial realization of Φ.*

6.4 A Kalman-Ho-like partial-realization algorithm for LSS

While Algorithm 1 is theoretically attractive because of its simplicity, it is not necessarily the most suitable one for numerical implementation. In Algorithm 2 we present an alternative algorithm for computing a semi-reachable and observable $2N+1$-partial realization of Φ. The algorithm is based on factorization of the Hankel-matrix $H_{\Phi,N+1,N}$ and it is similar to the Kalman-Ho algorithm.

The result of Algorithm 2 is described in the theorem below.

Algorithm 2 ComputePartialRealization($H_{\Phi,N+1,N}$)

1: Compute a decomposition of $H_{\Phi,N+1,N}$

$$H_{\Phi,N+1,N} = OR$$

where $O \in \mathbb{R}^{(Q^{\leq N+1} \times \{1,2,\ldots,pD\}) \times r}$ and $R \in \mathbb{R}^{r \times (Q^{\leq N} \times J_{\Phi})}$ are matrices such that $r = \text{rank } R = \text{rank } O = \text{rank } H_{\Phi,N+1,N}$.

2: Recall from (16) the enumeration of Q. For each $q \in Q$, $q = \sigma_i$, for some $i = 1, 2, \ldots, D$ define the matrix $\widetilde{C}_q \in \mathbb{R}^{p \times r}$ by

$$\widetilde{C}_q = \left[O^T_{(\epsilon,(i-1)p+1),.}, \; O^T_{(\epsilon,(i-1)p+2),.}, \; \ldots, \; O^T_{(\epsilon,ip),.} \right]^T$$

where $O_{k,.}$ denotes the row of O indexed by k.

3: For each $q \in Q$, define the matrix $\widetilde{B}_q \in \mathbb{R}^{r \times m}$ by

$$\widetilde{B}_q = \left[R_{.,(\epsilon,(q,1))}, \; R_{.,(\epsilon,(q,2))}, \; \cdots \; R_{.,(\epsilon,(q,m))} \right]$$

where $R_{.,(\epsilon,(q,j))}$ stands for the column of R indexed by $(\epsilon, (q, j))$ for $q \in Q$ and $j = 1, 2, \ldots, m$.

4: Define the map $\widetilde{\mu}_N : \Phi \to \mathbb{R}^r$ as

$$\forall f \in \Phi : \widetilde{\mu}_N(f) = R_{.,(\epsilon,f)}$$

where $R_{.,(\epsilon,f)}$ stands for the column of R indexed by (ϵ, f) for $f \in \Phi$.

5: For each $q \in Q$ let the matrix $\widetilde{A}_q \in \mathbb{R}^{r \times r}$ be the solution of equation

$$\bar{\Gamma}\widetilde{A}_q = \bar{\Gamma}_q \tag{26}$$

where $\bar{\Gamma}, \bar{\Gamma}_q \in \mathbb{R}^{(Q^{\leq N} \times \{1,2,\ldots,pD\}) \times r}$ are matrices of the form

$$\bar{\Gamma}_{(v,i),j} = O_{(v,i),j} \text{ and } (\bar{\Gamma}_q)_{(v,i),j} = O_{(qv,i),j}$$

for all $(v, i) \in Q^{\leq N} \times \{1, 2, \ldots, pD\}$, $j = 1, 2, \ldots, r$. That is, $\bar{\Gamma}$ is obtained from O by deleting all the rows indexed by pairs (v, i) where v is a sequence of discrete modes of length $N + 1$. The matrix $\bar{\Gamma}_q$ is the shifted version of O, i.e. its row indexed by (v, i) is the row of O indexed by (qv, i).

6: If there no unique solution to (26) then return *NoRealization*. Otherwise return $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ where $\widetilde{\Sigma}_N$ is a LSS of the form

$$\widetilde{\Sigma}_N = (\mathbb{R}^r, \mathcal{U}, \mathcal{Y}, \{(\widetilde{A}_q, \widetilde{B}_q, \widetilde{C}_q) \mid q \in Q\})$$

Theorem 5 (Partial realization algorithm) *With the notation above the following holds.*

(1) Assume that form some $N > 0$, (25) of Theorem 3 holds. Then Algorithm 2 always returns a LSS realization $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ and $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ is a minimal $2N+1$-realization of Φ. In fact, $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ is isomorphic to the LSS realization (Σ_N, μ_N) of Theorem 3.

(2) If for some $N > 0$, $\mathrm{rank}\ H_{\Phi,N,N} = \mathrm{rank}\ H_{\Phi}$, then (25) holds, and Algorithm 2 returns a minimal realization $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ of Φ. If there exists a LSS realization (Σ, μ) of Φ, such that $\dim \Sigma \leq N+1$, then $\mathrm{rank}\ H_{\Phi,N,N} = \mathrm{rank}\ H_{\Phi}$ holds and the realization $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ returned by Algorithm 2 is a minimal realization of Φ.

The proof of Theorem 5 is presented in Section 7.

Theorem 5 has immediate consequences for systems identification and model reduction of linear switched systems, by guaranteeing correctness of the realization algorithm described in Algorithm 2. In Section 8 we discuss the potential applications of Theorem 3 and Theorem 5 in more detail.

Remark 5 (Factorization of $H_{\Phi,N+1,N}$) *There are many algorithms to compute a factorization of $H_{\Phi,N+1,N}$ and Algorithm 2 can work with any of these algorithms. In particular, the factorization of $H_{\Phi,N+1,N}$ can be done through singular value decomposition. More precisely, assume that Φ is finite, and let $H_{\Phi,N+1,N} = U\Sigma V$ be the singular value decomposition of $H_{\Phi,N+1,N}$. Notice that if Φ is finite, then the index sets of $H_{\Phi,N+1,N}$ are finite and the singular value decomposition can be computed by enumerating the index sets of $H_{\Phi,N+1,N}$ and viewing $H_{\Phi,N+1,N}$ as a usual finite matrix. The choice $O = U\Sigma^{1/2}$ and $R = \Sigma^{1/2}V$ yields a decomposition $H_{\Phi,N+1,N} = OR$ satisfying the conditions of Algorithm 2. Algorithm 2 has been implemented, and the factorization used in the implementation is precisely the singular value decomposition of $H_{\Phi,N+1,N}$.*

7 Proof of the partial realization theorems for LSS

The goal of this section is to present the proof of Theorem 3, 4, and 5. Both proofs rely heavily on the relationship between rational representations and LSSs described in [45,42]. The outline of the section is the following. In Subsection 7.1 we will present a short overview of the relationship between rational representations and LSSs . Subsection 7.2 presents the proof of Theorem 3, 4, and 5. The reader is advised to review Appendix A and Appendix B before reading this section. Throughout the section we will tacitly use the notation and terminology of Appendix A and Appendix B.

Below we will present a brief overview of the relationship between formal power series and their representations and LSSs . For a more detailed presentation see [45,42]. Informally, the relationship is as follows.

- Let Φ be a family of input-output maps and assume that Φ has a generalized kernel representation. Then we can construct a family of formal power series Ψ_{Φ} from the generalized Markov-parameters of Φ. The details of the construction will be presented below. The family Ψ_{Φ} will be referred to as *the family of formal power series associated with* Φ.
- If (Σ,μ) is a LSS realization of Φ, then we can construct a rational representation $R_{\Sigma,\mu}$ of Ψ_{Φ} from the parameters of (Σ,μ). The converse is also true; if R is a representation of Ψ_{Φ}, then we can construct a realization (Σ_R,μ_R) of Φ from the parameters of R. The details of the construction will be presented below. We will call $R_{\Sigma,\mu}$ the *representation associated with* (Σ,μ) and we will call (Σ_R,μ_R) the *realization associated with* R.

We will start with defining the family of formal power series Ψ_{Φ} associated with Φ. To this end, recall from Definition 16 the definition of Markov-parameters $S_{q,q_0,j}(w)$, $S_{f,q}(w)$ of Φ, for each pair of discrete modes $q,q_0 \in Q$, input-output map $f \in \Phi$, index $j=1,2,\ldots,m$, and a word $w \in Q^*$. It is easy to see that the maps $S_{q,q_0,j} : Q^* \ni w \to S_{q,q_0,j}(w) \in \mathbb{R}^p$ and $S_{f,q} : Q^* \ni w \to S_{f,q}(w) \in \mathbb{R}^p$ define formal power series $S_{q,q_0,j}$ and $S_{f,q}$ in $\mathbb{R}^p \ll Q^* \gg$. Recall the enumeration of the set of discrete modes Q defined in (16). That is, Q assumed to have D elements given by the distinct elements $\sigma_1,\sigma_2,\ldots,\sigma_D$. For each discrete mode $q \in Q$, index $j=1,2,\ldots,m$, and input-output map $f \in \Phi$ define the formal power series $S_{q,j}, S_f \in \mathbb{R}^{pD} \ll Q^* \gg$ as follows; for each word $w \in Q^*$ let

$$\begin{aligned} S_{q,j}(w) &= \Big[(S_{\sigma_1,q,j}(w))^T,\,(S_{\sigma_2,q,j}(w))^T,\,\cdots,\,(S_{\sigma_D,q,j}(w))^T\Big]^T \in \mathbb{R}^{pD}, \\ S_f(w) &= \Big[(S_{f,\sigma_1}(w))^T,\,(S_{f,\sigma_2}(w))^T,\,\cdots,\,(S_{f,\sigma_D}(w))^T\Big]^T \in \mathbb{R}^{pD} \end{aligned} \tag{27}$$

That is, the values of the formal power series $S_{q,j}$ are obtained by stacking up the values of $S_{\sigma_i,q,j}$ for $i=1,2,\ldots,D$. Similarly, the values of S_f are obtained by stacking up the values of S_{f,σ_i} for $i=1,2,\ldots,D$. Define the set $J_{\Phi} = \Phi \cup \{(q,z) \mid q \in Q, z=1,2,\ldots,m\}$. Define the *indexed set of formal power series associated with* Φ as the following family of formal power series

$$\Psi_{\Phi} = \{S_j \in \mathbb{R}^{pD} \ll Q^* \gg \mid j \in J_{\Phi}\} \tag{28}$$

Remark 6 (Equivalence of definitions of the Hankel-matrix) *It is easy* *to see that the Hankel-matrix H_{Ψ_Φ} of the family of formal power series Ψ_Φ is identical to the Hankel-matrix H_Φ of Φ as defined in Definition 17*

Next we define the rational representation $R_{\Sigma,\mu}$ associated with a LSS realization (Σ, μ). Let Σ be a LSS of the form (5) and assume that $\mu : \Phi \to \mathcal{X}$ is a map assigning to each element $f \in \Phi$ an initial state of Σ.

Construction 1 (Representation associated with a realization) *Define the* representation associated with (Σ, μ) *as the rational $pD - J_\Phi$ representation*

$$R_{\Sigma,\mu} = (\mathcal{X}, \{A_q\}_{q\in Q}, \widetilde{B}, \widetilde{C})$$

The various components of $R_{\Sigma,\mu}$ are defined as follows.

- ***State-space $\mathcal{X}$.*** *The state-space of the rational representation $R_{\Sigma,\mu}$ is the same as the state-space of Σ, i.e. $\mathbb{R}^n = \mathcal{X}$.*
- ***Alphabet.*** *The representation $R_{\Sigma,\mu}$ is defined over the alphabet which equals the set of discrete modes Q.*
- ***State-transition maps (matrix)*** $\{A_q \in \mathbb{R}^n\}_{q\in Q}$***.*** *For each discrete mode $q \in Q$, the corresponding state-transition matrix A_q of $R_{\Sigma,\mu}$ is identical to the matrix A_q of Σ.*
- ***Readout matrix (map)*** $\widetilde{C} \in \mathbb{R}^{pD\times n}$***.*** *The readout matrix $\widetilde{C}$ is obtained by vertically "stacking up" the matrices $C_{\sigma_1}, \dots, C_{\sigma_D}$ in this order from top to bottom. That is, the $pD \times n$ matrix $\widetilde{C}$ is of the form*

$$\widetilde{C} = \begin{bmatrix} C_{\sigma_1}^T \ C_{\sigma_2}^T, \cdots \ C_{\sigma_D}^T \end{bmatrix}^T .$$

- ***Initial states****. The indexed set of the initial states of $R_{\Sigma,\mu}$ is of the form*

$$\widetilde{B} = \{\widetilde{B}_j \in \mathcal{X} \mid j \in J_\Phi\},$$

  i.e. it is indexed by the elements of the index set J_Φ. Its elements are defined by $\widetilde{B}_f = \mu(f)$ if f is an element of the family Φ, and $\widetilde{B}_{(q,l)} = (B_q)_{.,l}$, i.e. $\widetilde{B}_{(q,l)}$ is the lth column of B_q, for all $q \in Q$ and $l = 1, 2, \dots m$.

Conversely, below we will construct a LSS realization (Σ_R, μ_R) from any rational representation R satisfying some mild conditions. We would like to note that these conditions are automatically satisfied by a suitably chosen isomorphic copy of any rational representation of Ψ_Φ. The definition goes as follows.

Construction 2 (Realization associated with a representation) *Consider a $pD - J_\Phi$ representation R of the following form*

$$R = (\mathcal{X}, \{A_q\}_{q\in Q}, \widetilde{B}, \widetilde{C}) \qquad (29)$$

and assume that the following holds.

(1) The state-space is of the form $\mathcal{X} = \mathbb{R}^n$ for some $n > 0$. [2]
(2) The readout map $\widetilde{C}$ takes its values in $\mathbb{R}^{pD}$.
(3) The set of initial states $\widetilde{B}$ of R is indexed by the index set $J_\Phi = \Phi \cup \{(q, j) \mid q \in Q, j = 1, 2, \ldots, m\}$.

Define the LSS realization (Σ_R, μ_R) associated with R *as follows. Let Σ_R be of the form (5), where*

- ***State-space.*** *The state-space $\mathbb{R}^n$ of Σ_R is the same as that of R.*
- ***System matrices*** $\{A_q \in \mathbb{R}^{n\times n}\}_{q\in Q}$ ***of*** Σ_R. *For each discrete mode $q \in Q$, the matrix A_q of Σ_R is identical to the state-transition matrix A_q of R.*
- ***System matrices*** $\{C_q \in \mathbb{R}^{p\times n}\}_{q\in Q}$ ***of*** Σ_R. *For any discrete state $q \in Q$ of the form $q = \sigma_i$ for some $i = 1, 2, \ldots, D$ and for any $l = 1, 2, \ldots, p$, the lth row of C_q equals the $p(i-1)+l$-th row of $\widetilde{C}$, i.e.*

$$\widetilde{C} = \left[C_{\sigma_1}^T,\, C_{\sigma_2}^T,\, \cdots\, C_{\sigma_N}^T\right]^T.$$

- ***System matrices*** $\{B_q \in \mathbb{R}^{n\times m}\}_{q\in Q}$ ***of*** Σ_R. *For each discrete mode $q \in Q$ the $n \times m$ matrix B_q is obtained as follows; the lth column of B_q equals the initial state $\widetilde{B}_{q,l}$ for all $l = 1, 2, \ldots, m$. That is*

$$B_q = \left[\widetilde{B}_{(q,1)},\, \widetilde{B}_{(q,2)},\, \cdots,\, \widetilde{B}_{(q,m)}\right]$$

The map $\mu_R : \Phi \to \mathcal{X}$ assigns to each element f of Φ the initial state of R indexed by f, i.e.

$$\mu_R(f) = \widetilde{B}_f \text{ for all } f \in \Phi.$$

The following theorem states the relationship between representations and realizations formally.

Remark 7 *If we apply Construction 1 to (Σ_R, μ_R) then the resulting representation R_{Σ_R,μ_R} coincides with R, i.e. $R_{\Sigma_R,\mu_R} = R$. Conversely, if we apply Construction 2, to the representation $R_{\Sigma,\mu}$ associated with a LSS realization (Σ, μ), then we get (Σ, μ) back, i.e. $\Sigma_{R_{\Sigma,\mu}} = \Sigma$, $\mu_{R_{\Sigma,\mu}} = \mu$.*

Theorem 6 ([44,42,45]) *Let Φ be a family of input-output maps and assume that Φ admits a generalized kernel representation. With the notation and assumptions above the following holds.*

- *The realization (Σ, μ) is a LSS realization of Φ if and only if $R_{\Sigma,\mu}$ is a rational representation of Ψ_Φ.*

[2] If $\mathcal{X} = \mathbb{R}^n$ does not hold, then replace R with the isomorphic copy $\mathcal{S}R$ defined in (A.7), Remark 9 whose state-space is $\mathbb{R}^n$. Since $\mathcal{S}R$ and R are isomorphic, if R is a representation of Ψ_Φ, then $\mathcal{S}R$ will be a representation of Ψ_Φ as well.

- *The representation R is a representation of Ψ_Φ if and only if (Σ_R, μ_R) is a realization of Φ.*
- *The realization (Σ, μ) is semi-reachable if and only if $R_{\Sigma,\mu}$ is reachable, and (Σ, μ) is observable if and only if $R_{\Sigma,\mu}$ is observable. Similarly, (Σ_R, μ_R) is semi-reachable if and only if R is reachable, and (Σ_R, μ_R) is observable if and only if R is observable.*
- *If (Σ, μ) is a minimal realization of Φ, then $R_{\Sigma,\mu}$ is a minimal representation of Ψ_Φ. Conversely, if R is a minimal representation of Ψ_Φ, then (Σ_R, μ_R) is a minimal realization of Φ.*
- *The map $\mathcal{S} : (\Sigma, \mu) \to (\Sigma^{'}, \mu^{'})$ is a LSS morphism if and only if $\mathcal{S} : R_{\Sigma,\mu} \to R_{\Sigma^{'},\mu^{'}}$ is a representation morphism.*[3] *In addition, S is a representation isomorphism if and only if S is a LSS isomorphism. Conversely, if the representations R and $R^{'}$ satisfy the assumptions of Construction 2, then $\mathcal{S} : R \to R^{'}$ is a representation morphism if and only if $\mathcal{S} : (\Sigma_R, \mu_R) \to (\Sigma_{R^{'}}, \mu_{R^{'}})$ is a LSS morphism. Moreover, in this case $\mathcal{S}$ is a representation isomorphism if and only if it is a LSS isomorphism.*

7.2 Proofs of the results on partial realization of LSSs

The goal of this section is to present the proof of Theorem 3, Theorem 4 and Theorem 5. Before proceeding to the proof, we will need the following result on the relationship between the partial-realization problem for LSSs and the partial-realization problem for rational representations.

Theorem 7 *Assume that Φ is a family of input-output maps admitting a generalized kernel representation. A LSS realization (Σ, μ) is an $2N+1$-partial realization of Φ if and only of the associated representation $R_{\Sigma,\mu}$ is a $2N+1$-partial representation of the family of formal power series Ψ_Φ. Conversely, a representation R is a $2N+1$-partial representation of the family Ψ_Φ of formal power series associated with Φ, if and only if the LSS realization (Σ_R, μ_R) associated with R is a $2N+1$-partial realization of Φ.*

PROOF. The second statement of the theorem follows from the first one by noticing that $R = R_{\Sigma_R,\mu_R}$. Hence, it is enough to prove the first statement of the theorem. It follows easily from Definition 18 and the definition of the formal power series $S_f, S_{q_0,j}$, $f \in \Phi$, $q_0 \in Q$, $j = 1, 2, \ldots, m$, that (Σ, μ) is an $2N + 1$-partial realization of Φ, if and only if for all $w \in Q^{\leq 2N+1}$, $q_0 \in Q$,

[3] Notice that the state-space of (Σ, μ) coincides with that of $R_{\Sigma,\mu}$ and the state-space of $(\Sigma^{'}, \mu^{'})$ coincides with that of $R_{\Sigma^{'},\mu^{'}}$. Hence S can indeed be viewed as a representation morphism, provided that it commutes with the matrices of the representations involved

$f \in \Phi$, $j = 1, 2, \ldots, m$

$$\begin{aligned} S_{q_0,j}(w) &= \begin{bmatrix} C_{\sigma_1}^T \ C_{\sigma_2}^T \ \cdots \ C_{\sigma_D}^T \end{bmatrix}^T A_w B_{q_0} e_j \\ S_f(w) &= \begin{bmatrix} C_{\sigma_1}^T \ C_{\sigma_2}^T \ \cdots \ C_{\sigma_D}^T \end{bmatrix}^T A_w \mu(f) \end{aligned} \tag{30}$$

From the definition of $R_{\Sigma,\mu}$ it follows directly that (30) is equivalent to $R_{\Sigma,\mu}$ being an $2N+1$-representation of Ψ_Φ.

Now we are ready to present the proof of Theorem 3.

PROOF. [Proof of Theorem 3] First, recall from Definition 29 the definition of the matrix $H_{\Phi_\Psi,K,L}$ for some $K, L > 0$. It is easy to see that the matrix $H_{\Phi,K,L}$ as defined in Definition 20 and the matrix $H_{\Phi_\Psi,K,L}$ coincide, in particular rank $H_{\Phi,K,L} = \text{rank } H_{\Psi_\Phi,K,L}$. Recall from Remark 6 that the matrices H_Φ and H_{Φ_Ψ} coincide and hence their ranks are equal as well. We will prove the two statements of the theorem separately.

Proof of existence and computability of a partial realization
We will apply Theorem 10 to Ψ_Φ. Using the remark above, the condition (25) can be rewritten as (B.2) of Theorem 10. Consider the $pD - J_\Phi$ representation $R_N = (\text{Im} H_{\Phi,N,N+1}, \{A_q\}_{q\in Q}, B, C)$ defined in Theorem 10. Since (B.2) holds, we get that R_N is well-defined and it is a $2N+1$-partial representation of Ψ_Φ, and R_N is reachable and observable. Consider Algorithm 1 and the isomorphism $\mathcal{S}$ defined there. It is clear that $\mathcal{S}$ maps the column space of $H_{\Psi_\Phi,N,N+1} = H_{\Phi,N,N+1}$ to $\mathbb{R}^n$ with $n = \text{rank } H_{\Phi,N,N+1}$. Consider the isomorphic copy $\mathcal{S}R_N$ of R_N via the morphism $\mathcal{S}$, i.e. $\mathcal{S}R_N = (\mathbb{R}^n, \{\mathcal{S}A_q\mathcal{S}^{-1}\}_{q\in Q}, \{\mathcal{S}(B_j) \mid j \in J_\Phi\}, C\mathcal{S}^{-1})$. Obviously, $\mathcal{S}R_N$ is also a $2N+1$-partial representation of Ψ_Φ, and it is reachable and observable.

Consider now the LSS realization $(\Sigma_{\mathcal{S}R_N}, \mu_{\mathcal{S}R_N})$ associated with $\mathcal{S}R_N$, as defined by Construction 2.

Notice that if (25) holds, then (23) of Algorithm 1 has a unique solution. Indeed, (23) can be rewritten as $A_q R = R_q$. Here $R, R_q \in \mathbb{R}^{n\times(Q^{\leq N}\times J_\Phi)}$ and for each $(w, j) \in Q^{\leq N} \times J_\Phi$, the column of R indexed by (w, j) is the image $\mathcal{S}(\mathbf{C}_{w,j})$ of the column of $H_{\Phi,N,N}$ indexed by (w, j); the column of R_q indexed by (w, j) is the image $\mathcal{S}(\mathbf{C}_{wq,j})$ of the column of $H_{\Phi,N,N+1}$ indexed by (wq, j). The column space of R is the column space of $H_{\Phi,N,N}$ by $\mathcal{S}$, hence if (25) holds, then rank $R = n$. That is, the solution of $A_q R = R_q$, if it exists, is unique.

Hence, if (25) holds, then the LSS realization (Σ_N, μ_N) is uniquely defined by (23),(22),(21) and (24), if it exists. It is easy to see that the LSS realization $(\Sigma_{\mathcal{S}R_N}, \mu_{\mathcal{S}R_N})$ satisfies (23),(22),(21) and (24) and hence it coincides with

(Σ_N, μ_N). In other words, the LSS realization (Σ_N, μ_N) returned by Algorithm 1 equals the LSS realization $(\Sigma_{\mathcal{S}R_N}, \mu_{\mathcal{S}R_N})$ associated with the isomorphic copy $\mathcal{S}R_N$ of R_N, i.e. $(\Sigma_{\mathcal{S}R_N}, \mu_{\mathcal{S}R_N}) = (\Sigma_N, \mu_N)$. By Theorem 7 it follows then that (Σ_N, μ_N) is a $2N+1$-partial realization of Φ. In addition, from Theorem 6 it follows that (Σ_N, μ_N) is semi-reachable and observable.

Proof of existence of a complete realization
In addition, from Theorem 10 it follows that if rank $H_{\Phi,N,N} =$ rank H_Φ, i.e. rank $H_{\Psi_\Phi,N,N} =$ rank H_{Ψ_Φ}, then R_N is a minimal representation of Ψ_Φ, and hence so is $\mathcal{S}R_N$. But then by Theorem 6, $(\Sigma_{\mathcal{S}R_N}, \mu_{\mathcal{S}R_N}) = (\Sigma_N, \mu_N)$ is a minimal realization of Ψ_Φ. Finally, assume that there exists a realization (Σ, μ) of Φ, such that $\dim \Sigma \leq N$. Consider the representation $R = R_{\Sigma,\mu}$ associated with (Σ, μ) as defined in Construction 1. Notice that $\dim R = \dim \Sigma \leq N$. Then by Theorem 6, R is a representation of Ψ_Φ. Applying Theorem 10 we get that rank $H_{\Psi_\Phi} =$ rank $H_{\Psi_\Phi,N,N}$ and hence rank $H_\Phi =$ rank $H_{\Phi,N,N}$ holds, and that R_N, and hence $\mathcal{S}R_N$, is a minimal representation of Ψ_Φ. Again, from Theorem 6 it follows then that $(\Sigma_N, \mu_N) = (\Sigma_{\mathcal{S}R_N}, \mu_{\mathcal{S}R_N})$ is a minimal realization of Φ.

Next, we will present the proof of Theorem 4

PROOF. [Proof of Theorem 4] We will prove the statements of Theorem 4 one by one. Notice that if Φ satisfies (25), then the associated family of formal power series Ψ_Φ satisfies (B.2). This allows us to use Theorem 11 and Theorem 7 to prove the statements of the theorem.

Proof of Part 1
From the proof of Theorem 3 it follows that the LSS realization (Σ_N, μ_N) is the realization associated with an isomorphic copy of the representation R_N. By the remark above the family Ψ_Φ associated with Φ satisfies the condition of Theorem 11 and hence by Theorem 11 the representation R_N is a minimal $2N+1$-partial representation of Ψ_Φ. Then Theorem 7 implies that (Σ_N, μ_N) is a N-partial realization of Φ and its dimension is $\dim R_N$. It is easy to see that (Σ_N, μ_N) is a minimal $2N+1$-partial realization of Φ. For assume that (Σ, μ) is a $2N+1$-partial realization of Φ. Then by Theorem 7 the associated representation $R_{\Sigma,\mu}$ is a $2N+1$-partial representation of Ψ_Φ, and hence, $\dim \Sigma_N = \dim R_N \leq \dim R_{\Sigma,\mu} = \dim \Sigma$.

Proof of Part 2
If (Σ, μ) is a minimal $2N+1$-partial realization of Φ, then from Theorem 7 it follows that $R_{\Sigma,\mu}$ is a minimal $2N+1$-partial representation of Ψ_Φ. Indeed, consider a $2N+1$-partial representation R of Ψ_Φ. Using Remark 9, we can replace R by a suitable isomorphic copy, state-space of which is $\mathbb{R}^n$ with $n =$

$\dim R$. This isomorphic copy will also be a $2N+1$-partial representation of Ψ_Φ. Hence, without loss of generality we can assume that the state-space of R is $\mathbb{R}^n$ for $n = \dim R$ and hence Construction 2 can be applied. Then it follows from Theorem 7 that (Σ_R, μ_R) is a $2N+1$-partial realization of Φ. Hence, $\dim R_{\Sigma,\mu} = \dim \Sigma \leq \dim \Sigma_R = \dim R$ by minimality of (Σ, μ).

But then by Theorem 11 the representation $R_{\Sigma,\mu}$ has to be reachable, observable, and of dimension rank $H_{\Psi_\Phi,N,N} =$ rank $H_{\Phi,N,N}$. By Theorem 6, the latter means that (Σ, μ) is semi-reachable, observable, and of dimension rank $H_{\Phi,N,N}$.

Proof of Part 3
Let (Σ, μ) and $(\hat{\Sigma}, \hat{\mu})$ be two minimal $2N+1$-partial realizations of Φ. Using an argument analogous to the one presented in the proof of Part 2, one can show that the associated representations $R_{\Sigma,\mu}$ and $R_{\hat{\Sigma},\hat{\mu}}$ are minimal $2N+1$-partial representations of the family Ψ_Φ of formal power series associated with Φ. From Theorem 11 it follows that $R_{\Sigma,\mu}$ and $R_{\hat{\Sigma},\hat{\mu}}$ are isomorphic. From Theorem 6 it then follows that (Σ, μ) and $(\hat{\Sigma}, \hat{\mu})$ are isomorphic.

We will continue with the proof of Theorem 5

PROOF. [Proof of Theorem 5] The proof of the theorem relies on the following observation. Assume that Algorithm 3 is applied to Ψ_Φ. If we use the same factorization algorithm for factorizing the matrix $H_{\Phi,N+1,N} = H_{\Psi_\Phi,N+1,N}$ in both Algorithm 2 and Algorithm 3, then the following holds. The equation (26) has a unique solution if and only if (B.10) has a unique solution. In addition, Algorithm 2 returns a LSS realization if and only if Algorithm 3 applied to $H_{\Phi,N+1,N} = H_{\Psi_\Phi,N+1,N}$ returns a rational representation $\widetilde{R}_N$, and the realization $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ returned by Algorithm 2 is the LSS realization associated with the representation $\widetilde{R}_N$, as defined in Construction 2. That is, $(\widetilde{\Sigma}_N, \widetilde{\mu}_N) = (\Sigma_{\widetilde{R}_N}, \mu_{\widetilde{R}_N})$.

Proof of Part 1
Recall that for all $L, M > 0$, the matrix $H_{\Phi,L,M}$ is identical to $H_{\Psi_\Phi,L,M}$, and that the Hankel-matrix H_Φ is identical to H_{Ψ_Φ}. Hence, by applying Theorem 12 to $H_{\Psi_\Phi,N+1,N}$ we get that if (25) holds, then Algorithm 3 returns a $2N+1$-partial representation $\widetilde{R}_N$ of Ψ_Φ. Then from the discussion above it follows that Algorithm 2 returns a realization $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ and $(\widetilde{\Sigma}_N, \widetilde{\mu}_N) = (\Sigma_{\widetilde{R}_N}, \mu_{\widetilde{R}_N})$. Then from Theorem 7 it follows that $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ is a $2N+1$-partial realization of Φ. Moreover, Theorem 12 implies that in this case $\widetilde{R}_N$ is isomorphic to the representation R_N from Theorem 10. Recall from the proof of Theorem 3 that the realization (Σ_N, μ_N) from Theorem 3 is the realization associated with the isomorphic copy $\mathcal{S}R_N$ of the representation R_N, as defined in Construction 2. Since R_N and $\widetilde{R}_N$ are isomorphic, the representations $\mathcal{S}R_N$ and $\widetilde{R}_N$ are

isomorphic as well. Then by Theorem 6 the realizations $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ and (Σ_N, μ_N) are isomorphic as well.

Proof of Part 2
If rank $H_\Phi =$ rank $H_{\Phi,N,N}$, then by Theorem 3, (25) holds and (Σ_N, μ_N) is a minimal realization of Φ. Similarly, if there exists a LSS realization (Σ, μ) of Φ such that $\dim \Sigma \leq N$, then by Theorem 3, rank $H_\Phi =$ rank $H_{\Phi,N,N}$, and (25) holds, and (Σ_N, μ_N) is a minimal realization of Φ. By the preceding argument, in both cases Algorithm 2 returns a realization $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ and $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ is isomorphic to (Σ_N, μ_N). Hence, in both cases the realization $(\widetilde{\Sigma}_N, \widetilde{\mu}_N)$ returned by Algorithm 2 is a minimal realization of Φ.

8 Discussions and Conclusions

In this paper we presented basic results on partial-realization theory LSSs. We also discussed the algorithmic aspects of the theory and presented two algorithms for computing a partial LSS realization. We have also shown that under suitable conditions, the obtained partial realizations are in fact (complete) minimal realizations of the input-output maps at hand.

The paper represents only the first steps towards partial-realization theory of hybrid systems. A number of problems remains open. In particular, we would like to extend the results of the current paper to other classes of hybrid systems. Another important research direction is the improvement of the computational complexity of the presented algorithms.

As we mentioned it earlier, the results of the paper can potentially be useful for the solution of a number of problems, namely, for system identification, model reduction and possibly fault detection and computer vision. Below we will discuss these potential application domains in more detail.

Systems identification algorithms
As the history of development of system identification demonstrates, partial-realization theory can be used to design algorithms for system identification. In particular, the famous subspace identification algorithms [11,53] rely heavily on partial-realization theory.

In fact, based on the results of the paper one can derive the following crude identification algorithm. Consider a finite family $\Phi = \{f_1, f_2, \ldots, f_d\}$ of input-output maps. We would like to find a LSS realization of Φ. Fix a natural number N. The number N may be chosen either based on our belief on the dimension of a potential LSS realization of Φ, or our ability/willingness to make measurements of the system responses for switching sequences up to length $2N+1$.

1: For each $f \in \Phi$, estimate the first-order time derivatives of $D^{(1,1,1,\ldots,1,0)} f_{0,wq}$ and $D^{(1,1,1,\ldots,1,0)} f_{e_j,q_0wq} - D^{(1,1,1,\ldots,1,0)} f_{0,q_0wq}$ for all sequences of discrete modes w of length at most $2N+1$ and all discrete modes $q_0, q \in Q$. Use these estimates to construct the Markov-parameters of order up to $2N+1$. Subsequently, use the thus obtained Markov-parameters to construct the Hankel-matrices $H_{\Phi,N,N}$, $H_{\Phi,N+1,N}$ and $H_{\Phi,N,N+1}$.

2: If rank $H_{\Phi,N,N} =$ rank $H_{\Phi,N+1,N} =$ rank $H_{\Phi,N,N+1}$, then compute the $2N+1$ partial realization (Σ_N, μ_N) either by Algorithm 1 or Algorithm 2.

It follows immediately from Theorem 3 and Theorem 5 that if Φ can indeed be realized by a LSS , then for large enough N the procedure above will return a minimal LSS realization of Φ. In fact, if $N+1$ is the dimension of a potential LSS realization of Φ, then the algorithm above always returns a minimal LSS realization of Φ.

The algorithm described above has several drawbacks. For example, it is unclear how to obtain in practice the derivatives needed to compute the generalized Markov-parameters. Another problem that the algorithm does not take into account measurement noise. The authors are hopeful that these shortcoming can be overcome, as similar problems were successfully dealt with in the case of linear systems. However, we would like to remark that turning partial-realization theory to practically usable identification algorithms might take considerable amount of time; in case of linear systems it took several decades.

Spaces of LSSs

Partial realization theory can potentially be useful for studying the geometry and topology of the space of LSSs. In turn, the latter is useful for parametric system identification, as it helps to formalize the well-known paradigm of system identification: "find a model of a certain class which matches the measured data best". In addition, the geometric insights and the resulting notions of distance between systems can be useful for fault detection and computer vision [37,8,9,54,46]. For bilinear systems, the first steps were made in [48]. For linear systems, the topology and geometry of Hankel-matrices were investigated in [16,3,36,25,24,26,27]. Note that spaces of linear systems were also studied using a different approach, namely, by looking at equivalence classes of minimal linear systems under algebraic similarity. This approach resulted in deep results and insights in the geometry of spaces of systems and important algorithms for parametric identification, see [23,22,21,19,20,38,4,5].

Model reduction

The presented results on partial-realization theory can also be used for model reduction. Recall from [17,18,1] the moment matching approach to model reduction of linear systems. The core of this approach is to approximate a high-order to system with a lower order one, such that the lower order system is a a partial realization of a finite number of Markov-parameter of the original system. For LSSs the role of Markov-parameters (or moments) is played by the generalized Markov-parameters. Hence, one could try to extend the moment matching approach to LSSs and replace a high-order system with a lower order one which is a partial-realization of a finite sequence of generalized Markov-parameters. Of course, a great number of issues needs to be solved. In particular, obtaining good numerical algorithms will be a challenge.

Extension to more general classes of hybrid systems
We believe that partial-realization theory for LSSs can be useful for obtaining realization theory, system identification and model reduction for piecewise-affine hybrid systems with guards. One reason for this is that a piecewise-affine hybrid system with guards can be viewed as a feedback interconnection of a LSS with an event generator. The former is just the collection of the continuous subsystems of the piecewise-affine hybrid system, the latter is constructed from the guards of the piecewise-affine hybrid system. Understanding the realization theory of one of the components of this feedback loop should help to understand the realization theory of piecewise-affine hybrid systems. The second reason is that LSSs form a subclass of piecewise-affine hybrid systems. by identifying each LSSs with a piecewise-affine hybrid system whose guards depend only on the choice of inputs. Hence, any result on (partial-) realization theory of piecewise-affine hybrid systems should be consistent with the corresponding results for LSSs.

Acknowledgment The authors thank Pieter Collins and Luc Habets for useful discussions and suggestions.

References

[1] Athanasios C. Antoulas and Dan C. Sorensen. Approximation of large-scale dynamical systems: an overview. *Int. J. Appl. Math. Comput. Sci.*, 11(5):1093–1121, 2001. Numerical analysis and systems theory (Perpignan, 2000).

[2] J. Berstel and C. Reutenauer. *Rational series and Their Languages*. Springer-Verlag, 1984.

[3] Roger W. Brockett. The geometry of the partial realization problem. *Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on*, 17:1048–1052, Jan. 1978.

[4] Christopher I. Byrnes and Norman E. Hurt. On the moduli of linear dynamical systems. 4:83–122, 1979.

[5] Christopher I. Byrnes and Anders Lindquist. The stability and instability of partial realizations. *Systems Control Lett.*, 2(2):99–105, 1982/83.

[6] Frank M. Callier and Charles A. Desoer. *Linear System Theory*. Springer-Verlag, 1991.

[7] H. Chen and J.M. Maciejowski. New subspace identification method for bilinea systems. Technical Report CUED/FINFENG/TR.337, Department of Engineering, University of Cambridge, 2000.

[8] K. De Cock and B. De Moor. Subspace angles and distances between ARMA models. In *Proc. of the Int. Symposium of Math. Theory of Networks and Systems*, 2000.

[9] Gianfranco Doretto, Alessandro Chiuso, Ying Nian Wu, and Stefano Soatto. Dynamic textures. *International Journal of Computer Vision*, 51(2):91–109, 2003.

[10] Samuel Eilenberg. *Automata, Languages and Machines*. Academic Press, New York, London, 1974.

[11] W. Favoreel, B. De Moor, and P. Van Overschee. Subspace identification of bilinear systems subject to white inputs. Technical Report TR 1996-531, ESAT, Katholieke Univesiteit Leuven, 1996.

[12] M. Fliess. Matrices de hankel. *J. Math. Pures Appl.*, (23):197 – 224, 1973.

[13] M. Fliess. Functionnelles causales non linéaires et indéterminées non commutatives. *Bull. Soc. Math. France*, (109):2 – 40, 1981.

[14] F. Gécseg and I Peák. *Algebraic theory of automata*. Akadémiai Kiadó, Budapest, 1972.

[15] Dieter Gollmann. Partial realization by discrete-time internally bilinear systems: An algorithm. In *Mathematical theory of networks and systems, Proc. int. Symp., Beer Sheva/Isr. 1983*, 1983.

[16] William B. Gragg and Anders Lindquist. On the partial realization problem. *Linear Algebra Appl.*, 50:277–319, 1983.

[17] E. J. Grimme. *Krylov Projection Methods*. PhD thesis, Univ. Illinois, Urbana Champaign, 1997.

[18] E. J. Grimme, D. C. Sorensen, and P. Van Dooren. Model reduction of state space systems via an implicitly restarted Lanczos method. *Numer. Algorithms*, 12(1-2):1–31, 1996.

[19] B. Hanzon. Riemannian geometry on families of linear systems, the deterministic case. Technical Report 88-62, Delft University of Technology, Faculty of Mathematics and Informatics, 1988.

[20] B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, Part I and Part II*, volume 63 of *CWI Tract*. CWI, Amsterdam, 1989.

[21] B. Hanzon. On the differentiable manifold of fixed order stable linear systems. *Systems and Control Letters*, 13:345 – 352, 1989.

[22] M. Hazewinkel. Moduli and canonical forms for linear dynamical systems III: The algebraic-geometric case. In C. Martin and R. Hermann, editors, *The 1976 AMES Research Center (NASA) Conference on Geometric Control Theory*, pages 229–276. Brookline, Mass.: Math Sci Press 1977.

[23] M. Hazewinkel. Moduli and canonical forms for linear dynamical systems II: The topological case. *Mathematical Systems Theory*, 10:363–385, 1977.

[24] U. Helmke and D. Hinrichsen. Canonical forms and orbit spaces of linear systems. *IMA J Math Control Info*, 3:167 – 184, 1986.

[25] U. Helmke, D. Hinrichsen, and W. Manthey. A cell decomposition of the space of real Hankel matrices of rank $\leq n$ and some applications. *Linear Algebra Appl.*, 122/123/124:331–355, 1989.

[26] D. Hinrichsen, W. Manthey, and D. Prätzel-Wolters. The Bruhat decomposition of finite Hankel matrices. *Systems Control Lett.*, 7(3):173–182, 1986.

[27] Diederich Hinrichsen and Wilfried Manthey. On a cell decomposition for Hankel matrices and rational functions. *J. Reine Angew. Math.*, 451:15–50, 1994.

[28] B. L. Ho and R. E. Kalman. Effective construction of linear state-variable models from input/output data. In *Proc. Third Annual Allerton Conf. on Circuit and System Theory*, pages 449–459. Univ. Illinois, Urbana, Ill., 1965.

[29] A. Isidori. Direct construction of minimal bilinear realizations from nonlinear input-output maps. *IEEE Transactions on Automatic Control*, pages 626–631, 1973.

[30] Alberto Isidori. *Nonlinear Control Systems*. Springer Verlag, 1989.

[31] Alberto Isidori, Paolo D'Alessandro, and Antonio Ruberti. Realization and structure theory of bilinear dynamical systems. *SIAM J. Control*, 12(3), 1974.

[32] Nathan Jacobson. *Lectures in Abstract Algebra*, volume II: linear algebra. D. van Nostrand Company, Inc. New York, 1953.

[33] R. E. Kalman, P. L. Falb, and M. A. Arbib. *Topics in mathematical system theory*. McGraw-Hill Book Co., New York, 1969.

[34] W. Kuich and A. Salomaa. *Semirings, Automata, Languages*. Springer-Verlag, 1986.

[35] Daniel Liberzon. *Switching in Systems and Control*. Birkhäuser, Boston, 2003.

[36] Wilfried Manthey and Uwe Helmke. Bruhat canonical form for linear systems. *Linear Algebra Appl.*, 425(2-3):261–282, 2007.

[37] R. Martin. A metric for ARMA processes. *IEEE Trans. on SIgnal Processing*, 48(4):1164–1170, 2000.

[38] Ralf Peeters. *System Identification Based on Riemannian Geometry: Theory and Algorithms*. PhD thesis, Free University, Amsterdam, 1994.

[39] Mihaly Petreczky. Realization theory for bilinear hybrid systems. In *In Proceedings 11th IEEE International Conference on Methods and Models in Automation and Robotics*, 2005.

[40] Mihaly Petreczky. Realization theory for bilinear switched systems. In *Proc. of 44th IEEE Conference on Decision and Control*, 2005.

[41] Mihaly Petreczky. Realization theory for linear and bilinear hybrid systems. Technical Report MAS-R0502, CWI, 2005.

[42] Mihaly Petreczky. Realization theory of linear and bilinear switched systems: A formal power series approach. Technical Report MAS-R0403, CWI, 2005. To appear in ESAIM Control, Optimization and Calculus of Variations, DOI 10.1051/cocv/2010014 and DOI 10.1051/cocv/2010015.

[43] Mihály Petreczky. Hybrid formal power series and their application to realization theory of hybrid systems. In *Proceedings 17th International Symposium on Mathematical Theory of Networks and Systems*, 2006.

[44] Mihaly Petreczky. *Realization Theory of Hybrid Systems*. PhD thesis, Vrije Universiteit, Amsterdam, 2006. Available at `http://www.cwi.nl/~mpetrec`.

[45] Mihaly Petreczky. Realization theory for linear switched systems: Formal power series approach. *Systems and Control Letters*, 56(9 – 10):588–595, 2007.

[46] Mihály Petreczky and René Vidal. Metrics and topology for nonlinear and hybrid systems. In *Hybrid Systems: Computation and Control*, pages 459–472, 2007.

[47] Mihaly Petreczky and Rene Vidal. Realization theory of stochastic jump-markov linear systems. In *Proceedings 46th IEEE Conference on Decision and Control*, 2007.

[48] D. Eduardo Sontag. A remark on bilinear systems and moduli spaces of instantons. *Systems and Control Letters*, 9(5):361–367, 1987.

[49] Eduardo D. Sontag. *Polynomial Response Maps*, volume 13 of *Lecture Notes in Control and Information Sciences*. Springer Verlag, 1979.

[50] Eduardo D. Sontag. Realization theory of discrete-time nonlinear systems: Part I – the bounded case. *IEEE Transaction on Circuits and Systems*, CAS-26(4), April 1979.

[51] Zhendong Sun and Shuzhi S. Ge. *Switched linear systems : control and design*. Springer, London, 2005.

[52] A. Tether. Construction of minimal linear state-variable models from finite input-output data. *IEEE Transactions Automatic Control*, 15(4):427– 436, 1970.

[53] P. van Overschee and B. De Moor. *Subspace Identification for Linear Systems*. Kluwer Academic Publishers, 1996.

[54] S. V. N. Vishwanathan, Alexander J. Smola, and René Vidal. Binet-cauchy kernels on dynamical systems and its application to the analysis of dynamic scenes. *International Journal of Computer Vision*, 73(1):95–119, 2007.

A Formal Power Series

The section recalls basic results on formal power series. The material of this section is based on the classical theory of formal power series, see [2,34]. However, a number of concepts and results are extensions of the standard ones to families of formal power series. The outline of the section is the following. In Subsection A.1 we present the basic concepts for formal power series. In Subsection A.2 we present the notion of a rational representation of a family of formal power series. In Subsection A.3 we state the main result on existence and minimality of a rational representation of a family of formal power series.

A.1 Formal Power Series: Definition and Basic Concepts

Let X be a finite set, which we will refer to as the alphabet. Recall from Section 4.0.3 the notion of a finite word over an alphabet and the related notation. A *formal power series* S with coefficients in $\mathbb{R}^p$ is a map

$$S : X^* \to \mathbb{R}^p$$

There are many ways to give an intuition for the definition of a formal power series. For the purposes of this paper, the most suitable one is to think of a formal power series as an output of a machine defined as follows. The machine reads symbols belonging to X from its input tape and writes elements of $\mathbb{R}^p$ onto its output tape. We denote by $\mathbb{R}^p \ll X^* \gg$ the set of all formal power series with coefficients in $\mathbb{R}^p$. The set of all formal power series over X with coefficients in $\mathbb{R}^p$ forms a vector space with respect to point-wise addition and multiplication. That is, if $\alpha, \beta \in \mathbb{R}$ are two scalars and $S, T \in \mathbb{R}^p \ll X^* \gg$ are two formal power series, then the linear combination $\alpha S + \beta T$ is defined as the formal power series assigning each word $w \in X^*$ the value $\alpha S(w) + \beta T(w)$.

In the sequel we will mostly be interested in *families of formal power series.*

Definition 21 (Familly of formal power series) *Let J be an arbitrary (possibly infinite) set. A* family of formal power series in $\mathbb{R}^p \ll X^* \gg$ indexed by *J is simply a collection $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ of formal power series from $\mathbb{R}^p \ll X^* \gg$ indexed by elements of J.*

Notice that the definition above does not require S_j, $j \in J$ to be all distinct formal power series, i.e. $S_l = S_j$ for some indices $j, l \in J$, $j \neq l$ is allowed. One can think of a family of formal power series as a family of input-output maps of the machine (discrete-time system) described above, realized from a set of initial states indexed by elements of J. We would like to point out that the notion of (family of) formal power series is a purely formal one, the interpretation of them as input-output maps is just one of the many possible interpretations.

A.2 Rational Representations and Rational Formal Power Series

Above we have defined the notion of a formal power series and a family of formal power series and we have related these notions to input-output maps of some systems. Below we will recall the notion of a *rational representation*, which can be thought of as a special subclass of these systems. We will also define morphisms between rational representations along with a notion of observability and reachability.

Definition 22 (Representation) *Let J be an arbitrary set and let $\mathbb{N} \ni p > 0$. A* rational representation of type p-J over the alphabet X *is a tuple*

$$R = (\mathcal{X}, \{A_\sigma\}_{\sigma \in X}, B, C) \tag{A.1}$$

where

- *The space $\mathcal{X}$ is a finite dimensional vector space over $\mathbb{R}$, called* state-space *of R,*
- *For each letter $\sigma \in X$, $A_\sigma : \mathcal{X} \to \mathcal{X}$ is a linear map, referred to as the* state-transition map
- *The map $C : \mathcal{X} \to \mathbb{R}^p$ is a linear map, referred to as the* readout map*,*
- *The family $B = \{B_j \in \mathcal{X} \mid j \in J\}$ is a collection of (not necessarily distinct) elements of $\mathcal{X}$ indexed by J.*

If p and J are clear from the context, then we will refer to R simply as a rational representation*.*

The dimension $\dim \mathcal{X}$ *of the state-space is called the* dimension *of the representation R and it is denoted by* $\dim R$.

Remark 8 *Notice that if a basis of $\mathcal{X}$ is fixed and $n = \dim \mathcal{X}$, then the state-transition maps A_σ, $\sigma \in X$ and the readout map C can be identified with their matrix representations in this basis, and for each $j \in J$, B_j can be identified with the $\mathbb{R}^n$ column vector of its coordinates. If $\mathcal{X} = \mathbb{R}^n$, then we will identify the linear maps A_σ, $\sigma \in X$ and C with their matrix representations in the standard orthogonal basis of $\mathbb{R}^n$. In this case we will call them the* state-transition matrices *and the* readout matrix *respectively.*

Definition 23 (Rational family of formal power series) *Let* $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg | \ j \in J\}$ *be a family of formal power series indexed by* J*. The representation* R *from (A.1) is said to be a* representation of Ψ*, if for each index* $j \in J$*, for all sequences* $\sigma_1, \sigma_2, \ldots, \sigma_k \in X$*,* $k \geq 0$*,*

$$S_j(\sigma_1\sigma_2 \cdots \sigma_k) = CA_{\sigma_k}A_{\sigma_{k-1}} \cdots A_{\sigma_1}B_j. \quad \text{(A.2)}$$

We will say that a family of formal power series $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg | \ j \in J\}$ *is* rational*, if there exists a representation* R *such that* R *is a representation of* Ψ*.*

Notation 4 *The following notation will greatly simplify the expressions used for rational representations. Let* $A_\sigma : \mathcal{X} \to \mathcal{X}, \sigma \in X$ *be linear maps and let* $w = \sigma_1\sigma_2 \cdots \sigma_k \in X^*$*,* $\sigma_1, \sigma_2, \cdots \sigma_k \in X$*,* $k \geq 0$*, be a word over* X*. Then* A_w *denotes the composition of the linear maps* $A_{\sigma_1}, A_{\sigma_2}, \ldots A_{\sigma_k}$ *in that order, that is*

$$A_w = A_{\sigma_k}A_{\sigma_{k-1}} \cdots A_{\sigma_1} \quad \text{(A.3)}$$

If $w = \epsilon$ *is the empty words, then* A_ϵ *is taken to be the identity map.*

With the notation above, (A.2) can be rewritten as $S_j(w) = CA_wB_j$ for all $w \in X^*$, $j \in J$. Notice that if $\mathcal{X} = \mathbb{R}^n$ for some $\mathbb{N} \ni n > 0$, and hence the state-transition maps A_σ can be viewed as matrices, then (A.3) defines a notation for products of the state-transition matrices taken along the word w.

Definition 24 (Minimality) *A representation* R_{min} *of* Ψ *is called* minimal*, if for each representation* R *of* Ψ*,* $\dim R_{min} \leq \dim R$*, i.e.* R_{min} *is a rational representation of* Ψ *with the smallest possible state-space dimension.*

We will continue with presenting the notions of *observability* and *reachability* for rational representations. Define the subspaces W_R and O_R of $\mathcal{X}$ by

$$W_R = \text{Span}\{A_wB_j \in \mathcal{X} \mid w \in X^*, j \in J\} \quad \text{(A.4)}$$

$$O_R = \bigcap_{w \in X^*} \ker CA_w \quad \text{(A.5)}$$

That is, the subspace W_R is the linear span of the elements of the state-space of the form A_wB_j, where w runs through all words over X and j runs through all the indices in J. The space O_R is the intersection of the null-spaces (kernels) of all the linear maps $CA_w : \mathcal{X} \to \mathbb{R}^p$, where w runs through the set of all words X^*.

Definition 25 (Reachability) *We will say that the representation* R *is* reachable *if* $\dim W_R = \dim R$*. The subspace* W_R *will be referred to as the* reachability subspace of *R.*

Definition 26 (Observability) *We will say that* R *is* observable *if* $O_R = \{0\}$*, i.e.* O_R *consists of the zero element only. The subspace* O_R *will be referred to as the* observability subspace of R.

Next, we define the notion of morphism between rational representations.

Definition 27 (Representation morphism) *Let* $R = (\mathcal{X}, \{A_\sigma\}_{\sigma \in X}, B, C)$, $\widetilde{R} = (\widetilde{\mathcal{X}}, \{\widetilde{A}_\sigma\}_{\sigma \in X}, \widetilde{B}, \widetilde{C})$ *be two p-J rational representations. A linear map* $\mathcal{S} : \mathcal{X} \to \widetilde{\mathcal{X}}$ *is called a* representation morphism *from* R *to* $\widetilde{R}$ *and is denoted by* $\mathcal{S} : R \to \widetilde{R}$ *if* $\mathcal{S}$ *commutes with* A_σ, B_j *and* C *for all* $j \in J$, $\sigma \in X$*, that is, if the following equalities hold*

$$\mathcal{S}A_\sigma = \widetilde{A}_\sigma \mathcal{S}, \forall \sigma \in X, \, \mathcal{S}B_j = \widetilde{B}_j, \forall j \in J, \, C = \widetilde{C}\mathcal{S} \tag{A.6}$$

The representation morphism $\mathcal{S}$ *is called* surjective, injective, isomorphism *if* $\mathcal{S}$ *is a surjective, injective or isomorphism respectively if viewed as a linear map.*

Remark 9 *If* $R = (\mathcal{X}, \{A_\sigma\}_{\sigma \in X}, B, C)$ *is a representation of* Ψ*, then for any vector space isomorphism* $\mathcal{S} : \mathcal{X} \to \mathbb{R}^n$, $n = \dim R$*, the tuple*

$$\mathcal{S}R = (\mathbb{R}^n, \{\mathcal{S}A_\sigma \mathcal{S}^{-1}\}_{\sigma \in X}, \mathcal{S}B, C\mathcal{S}^{-1}) \tag{A.7}$$

where $\mathcal{S}B = \{\mathcal{S}B_j \in \mathbb{R}^n \mid j \in J\}$ *is also a representation of* Ψ*. Moreover, for all* $\sigma \in X$, $\mathcal{S}A_\sigma\mathcal{S}^{-1}$ *can be naturally viewed as a matrix by taking its matrix representation with respect to the natural basis of* $\mathbb{R}^n$*. Similarly, by taking matrix and vector representations of* $C\mathcal{S}^{-1}$ *and* $\mathcal{S}B_j$, $j \in J$ *in the natural basis of* $\mathbb{R}^n$*, we can view* $C\mathcal{S}^{-1}$ *and* $\mathcal{S}B_j$, $j \in J$ *as a* $p \times n$ *matrix and* $n \times 1$ *vector respectively. Moreover,* $\mathcal{S} : R \to \mathcal{S}R$ *is a representation isomorphism. That is, we can always replace a representation of* Ψ *with an isomorphic representation, state-space of which is* $\mathbb{R}^n$ *for some* n*, and the parameters of which are matrices and real vectors, as opposed to linear maps and elements of abstract vector spaces. Moreover, isomorphisms clearly preserve such properties as observability, reachability, and minimality.*

A.3 Existence and Minimality of Rational Representations: Main Results

The purpose of the section is to state the main results on existence and minimality of representations of families of rational formal power series.

We will start by stating the main result on the existence of a rational representation. However, in order to state the main theorem, we need to define the concept of the *Hankel matrix* of a family of formal power series. Let $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ be a family of formal power series. Recall the notation of Section 4.0.2.

Construction 3 (Hankel-matrix) *Define the* Hankel-matrix *of* Ψ *as the following infinite matrix* H_Ψ*. The rows of* H_Ψ *are indexed by pairs* (v, i) *where* $v \in X^*$ *is an arbitrary word and* $i = 1, \ldots, p$*. The columns of* H_Ψ *are indexed by pairs* (w, j) *where* $w \in X^*$ *is a word over* X *and* $j \in J$ *runs through the elements of* J*. The set of all real infinite matrices with the indexing the rows and columns described above will be denoted by* $\mathbb{R}^{(X^* \times I) \times (X^* \times J)}$ *where* $I = \{1, \ldots, p\}$*. Hence, we can write* $H_\Psi \in \mathbb{R}^{(X^* \times I) \times (X^* \times J)}$*. The entry of* H_Ψ *lying on the intersection of the row indexed by* (v, i) *and the column indexed by* (w, j) *is defined as*

$$(H_\Psi)_{(v,i)(w,j)} = (S_j(wv))_i \tag{A.8}$$

where $(S_j(wv))_i$ *denotes the ith entry of the column vector* $S_j(wv) \in \mathbb{R}^p$*.*

According to the convention adopted in Section 4.0.2, we define the *rank of* H_Ψ*, denoted by* rank H_Ψ, as the dimension of the vector space spanned by the columns of H_Ψ. With the notation above the following theorem holds.

Theorem 8 (Existence of a representation,[44,42]) *The family* Ψ *is rational, i.e. admits a rational representation, if and only if* rank $H_\Psi < +\infty$*, i.e. the rank of the Hankel-matrix* H_Ψ *is finite.*

The proof of the above theorem can be found in [44,42].

As the next step we will present below the main result on minimality of rational representations.

Theorem 9 (Minimal representation, [44,42]) *Let* $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ *be a family of formal power series. The following are equivalent.*

(i) R_{min} *is a minimal representation of* Ψ*.*
(ii) R_{min} *is reachable and observable.*
(iii) If R *is a reachable representation of* Ψ*, then there exists a surjective representation morphism* $\mathcal{S} : R \to R_{min}$*.*
(iv) rank $H_\Psi = \dim R_{min}$*.*

In addition, all minimal representations of Ψ *are isomorphic.*

The proof of the above theorem can be found in [44,42].

Remark 10 (Related work) *The counterpart of the above two theorems, i.e. Theorem 8 and 9, for a single formal power series is a classical result, see [12,2,34,49,50].*

B Partial-realization theory of formal power series

Below we will formulate and solve the counterpart of the partial realization problem for families of formal power series. We use the obtained result to derive a solution to the partial realization problem for linear switched systems formulated in Problem 2. We start with defining the notion of a N-partial representation.

Definition 28 (N-partial representation) *Let $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ be a family of formal power series indexed by J. A J-p representation $R = (\mathcal{X}, \{A_x\}_{x\in X}, C, B)$, with $B = \{B_j \in \mathcal{X} \mid j \in J\}$ is said to be an N-*partial representation of Ψ *if for each index $j \in J$ and each word $w \in X^*$ of length at most N, i.e. $|w| \le N$,*

$$S_j(w) = CA_wB_j$$

That is, if R is a N-partial representation of Ψ, then R recreates the values of the elements of Ψ for all the words of length at most N. Now we are ready to formulate the partial-realization problem for formal power series.

Problem 3 (Partial-realization problem for formal power series) *Let Ψ be a family of formal power series indexed by J.*

- *Find conditions for existence of a N-partial representation of Ψ. Formulate an algorithm for computing a N-partial representation of Ψ.*
- *Characterize minimal N-partial representations of Ψ, their existence and uniqueness.*
- *Find conditions under which the N-partial representation above becomes a rational representation of Ψ in the sense of Definition 22.*

We will devote the rest of the section to solving the problem described above. The outline of the section is the following. Subsection B.1 presents the main results on partial-realization theory of families of formal power series, along with a Kalman-Ho-like algorithm for computing a minimal partial representation. Subsection B.2 presents the proof of the results presented in Subsection B.1. Throughout the section, Ψ will denote the family of formal power series $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$.

B.1 Main results on partial-realization theory of formal power series

The goal of the section is to present the main results on partial-realization theory of families of formal power series. Note that partial-realization theory of a single formal power series is more or less equivalent to partial-realization theory of bilinear and state-affine systems. The latter was already investigated in [29,15,31,49,50,11,7]. The results to be presented below represent an extension of the results of [29,15,31,49,50,11,7]. In addition, here we state partial-realization theory directly for formal power series. This is in contrast to [29,15,31,49,50,11,7], where the partial-realization problem and solution were stated in terms of state-affine or bilinear systems.

The outline of the section is the following. Subsection B.1.1 presents the statement of the main results on partial realization theory for formal power series. Subsection B.1.2 presents an algorithm for computing a partial representation, which is similar to the well-known Kalman-Ho algorithm for the linear systems.

B.1.1 Partial realization theory

We will start with defining the upper-left block matrix of H_Ψ which is indexed by words over X of finite length. More precisely, fix natural numbers $M, K > 0$ and define the following sets

$$\begin{aligned} I_M &= \{(v,i) \mid v \in X^*, |v| \le M, i = 1,\dots,p\} \\ J_K &= \{(w,j) \mid j \in J, w \in X^*, |w| \le K\} \end{aligned} \tag{B.1}$$

Intuitively, I_M is a subset of the row indices of the Hankel-matrix H_Ψ, made up of indices of the form (v,i), where v runs through all the words of length at most M. Similarly, the set J_K is a subset of column indices of H_Ψ, made up of column indices (w,j) where w runs through all the words of length at most K. The set J_K is the set of columns indices of the matrix to be defined, and I_M is the set of row indices of this matrix.

Definition 29 (Sub-matrices of the Hankel-matrix H_Ψ) *Define the matrix $H_{\Psi,M,K} \in \mathbb{R}^{I_M \times J_K}$ by*

$$(H_{\Psi,M,K})_{(v,i),(w,j)} = (H_\Psi)_{(v,i),(w,j)} = (S_j(wv))_i$$

for all $(v,i) \in I_M$ and $(w,j) \in J_K$.

That is, $H_{\Psi,M,K}$ is the left upper $I_M \times J_K$ block matrix of H_Ψ. Notice that if J is finite, then $|J_K| < +\infty$, that is, $H_{\Psi,M,K}$ is a *finite matrix.*

It turns out that under certain circumstances partial representations not only exist but they also yield a minimal representation of the whole family of formal power series. Moreover, such partial representations can be constructed from finite data.

Theorem 10 (Existence of partial representation) *With the notation above the following holds.*

(1) If for some $N > 0$,

$$\text{rank } H_{\Psi,N,N} = \text{rank } H_{\Psi,N,N+1} = \text{rank } H_{\Psi,N+1,N} \tag{B.2}$$

then there exists a $2N+1$-partial representation R_N of Ψ, of the form

$$R_N = (\text{Im}H_{\Psi,N,N+1}, \{A_\sigma\}_{\sigma\in X}, B, C) \tag{B.3}$$

where the parameters of R_N are defined as follows. For each word $w \in X^*$, $|w| \leq N+1$, and each index $j \in J$ denote by $(H_{\Psi,N,N+1})_{.,(w,j)}$ the column of $H_{\Psi,N,N+1}$ indexed by (w,j). With this notation,

- $\text{Im}H_{\Psi,N,N+1}$ *denotes the linear space spanned by the columns of $H_{\Psi,N,N+1}$.*
- *For each $\sigma \in X$, the linear map $A_\sigma : \text{Im}H_{\Psi,N,N+1} \to \text{Im}H_{\Psi,N,N+1}$ has the property that for each word $w \in X^*$, $|w| \leq N$ and each index $j \in J$*

$$A_\sigma((H_{\Psi,N,N+1})_{.,(w,j)}) = (H_{\Psi,N,N+1})_{.,(w\sigma,j)} \tag{B.4}$$

  i.e. A_σ maps the column indexed by $(w,j) \in J_N$ to the column indexed by $(w\sigma, j)$.
- *For each $w \in X^*$, $|w| \leq N+1$ and $j \in J$, the linear map C : $\text{Im}H_{\Psi,N,N+1} \to \mathbb{R}^p$ satisfies,*

$$C((H_{\Psi,N,N+1})_{.,(w,j)}) = \Big[(H_{\Psi,N,N+1})_{(\epsilon,1),(w,j)},\ (H_{\Psi,N,N+1})_{(\epsilon,2),(w,j)},\ \cdots,\ (H_{\Psi,N,N+1})_{(\epsilon,p),(w,j)}\Big]^T \tag{B.5}$$

  That is, C maps each column to the vector in $\mathbb{R}^p$ formed by the first p entries associated with $v = \epsilon$ of the column indexed by (w,j).
- *The set $B = \{B_j \in \text{Im}H_{\Psi,N,N+1} \mid j \in J\}$ is defined by*

$$B_j = (H_{\Psi,N,N+1})_{.,(\epsilon,j)} \text{ for all } j \in J \tag{B.6}$$

  i.e. B_j is simply the column of $H_{\Psi,N,N+1}$ indexed by (ϵ, j).

In addition, the representation R_N is reachable and observable.

(2) If for some $N > 0$

$$\text{rank } H_{\Psi,N,N} = \text{rank } H_{\Psi} \tag{B.7}$$

then (B.2) holds and the representation R_N from (B.3) is a minimal representation of Ψ.

(3) If R is a representation of Ψ, $\dim R \leq N+1$, then (B.7) holds for N and the representation R_N from (B.3) exists and it is a minimal representation of Ψ.

The proof of Theorem 10 is presented in Appendix B.2. Using the results of Theorem 10 we can state the following characterization of minimal partial representations of Ψ.

Theorem 11 (Minimal partial representation) *With the notation of Theorem 10, if Ψ satisfies (B.2), then the following holds.*

(1) A minimal $2N+1$ partial representation of Ψ exists, in fact, the representation R_N of Theorem 10 is a minimal $2N+1$ partial representation of Ψ.
(2) Any minimal $2N+1$ partial representation of Ψ is reachable and observable and it is of dimension $\operatorname{rank} H_{\Psi,N,N}$.
(3) All minimal $2N+1$ partial representations of Ψ are isomorphic.

The proof of Theorem 11 is presented in Appendix B.2.

Remark 11 *The reader might wonder why we speak of $2N+1$-partial representations in Theorem 10 and 11. The reason behind it is that the finite Hankel-matrix $H_{\Psi,N,N+1}$ is formed by values of the formal power series from Ψ for words of length at most $2N+1$. We would like the representation obtained from $H_{\Psi,N,N+1}$ to recreate at least the entries of the matrix $H_{\Psi,N,N+1}$. But this means precisely that the representation obtained from $H_{\Psi,N,N+1}$ should be a $2N+1$-partial representation of Ψ.*

B.1.2 Partial realization algorithm

In this section we present an algorithm, described in Algorithm 3, which computes a partial representation by factorizing the Hankel-matrix. The technique of Hankel-matrix factorization has been used in realization theory and systems identification for several decades. It forms the theoretical basis of algorithms for subspace identification, see for example [11,7]. Throughout the section, Ψ stands for the family of formal power series $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$.

Algorithm 3 represents an algorithm based on matrix factorization of $H_{\Psi,N+1,N}$. In addition, if N is large enough, Algorithm 3 in fact yields a minimal representation of Ψ. Algorithm 3 above may return two different types of data. It returns a rational representation if (B.10) has a unique solution, and the symbol $NoRepresentation$ otherwise.

Remark 12 (Implementation of the matrix factorization) *In step 1 of Algorithm 3 above one can use any algorithm for computing a factorization. For example, one could use SVD decomposition, in which case $H_{\Psi,N+1,N} = U\Sigma V^T$, and $O = U(\Sigma^{1/2})$, $R = (\Sigma^{1/2})V^T$ is a valid choice for decomposition.*

The following theorem characterizes the outcome Algorithm 3.

Algorithm 3 ComputePartialRepresentation($H_{\Psi,N+1,N}$)

1: Compute a decomposition of $H_{\Psi,N+1,N}$

$$H_{\Psi,N+1,N} = OR$$

$O \in \mathbb{R}^{I_{N+1} \times r}$, $R \in \mathbb{R}^{r \times J_N}$, rank $R =$ rank $O =$ rank $H_{\Psi,N+1,N} = r$

2: Define the matrix $\widetilde{C} \in \mathbb{R}^{p \times r}$ by

$$\widetilde{C} = \left[O^T_{(\epsilon,1),.}, \; O^T_{(\epsilon,2),.}, \; \cdots \; O^T_{(\epsilon,p),.} \right]^T \tag{B.8}$$

where $O_{k,.}$ denotes the row of O indexed by $k \in I_{N+1}$.

3: Define the family of vectors $\widetilde{B} = \{\widetilde{B}_j \in \mathbb{R}^r \mid j \in J\}$, where for each $j \in J$,

$$\widetilde{B}_j = R_{.,(\epsilon,j)} \tag{B.9}$$

where $R_{.,(\epsilon,j)}$ stands for the column of R indexed by (ϵ, j).

4: For each $\sigma \in X$ let $\widetilde{A}_\sigma \in \mathbb{R}^{r \times r}$ be the solution of

$$\bar{\Gamma}\widetilde{A}_\sigma = \bar{\Gamma}_\sigma \tag{B.10}$$

where $\bar{\Gamma}, \bar{\Gamma}_\sigma \in \mathbb{R}^{I_N \times r}$ are matrices of the form

$$\bar{\Gamma}_{(u,i),j} = O_{(u,i),j} \text{ and } (\bar{\Gamma}_\sigma)_{(u,i),j} = O_{(\sigma u,i),j}$$

for all $(u, i) \in I_N, j = 1, 2, \dots, r$.

5: If there no unique solution to (B.10) then return *NoRepresentation*. Otherwise return

$$\widetilde{R}_N = (\mathbb{R}^r, \{\widetilde{A}_x\}_{x \in X}, \widetilde{B}, \widetilde{C})$$

Theorem 12 *Let $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ be a family of formal power series. With the notation above the following holds.*

(1) Assume that form some $N > 0$

$$\text{rank } H_{\Psi,N,N+1} = \text{rank } H_{\Psi,N+1,N} = \text{rank } H_{\Psi,N,N} \tag{B.11}$$

Then Algorithm 3 always returns a formal power series representation $\widetilde{R}_N$ and $\widetilde{R}_N$ is an $2N+1$-partial representation of Ψ. In fact, the representation R_N from Theorem 10 and $\widetilde{R}_N$ are isomorphic. Hence, $\widetilde{R}_N$ is a minimal $2N+1$-partial representation of Ψ, it is reachable and observable.

(2) If for some $N > 0$,

$$\text{rank } H_{\Psi,N,N} = \text{rank } H_\Psi \tag{B.12}$$

then (B.11) holds, and Algorithm 3 returns a minimal representation $\widetilde{R}_N$ of Ψ.

(3) Assume rank $H_\Psi \leq N+1$*, or , equivalently, there exists a representation R of Ψ, such that* $\dim R \leq N+1$*. Then (B.12) holds and the represen-*

tation $\widetilde{R}_N$ returned by Algorithm 3 is a minimal representation of Ψ.

Remark 13 (Solution to (B.10)) *Although solution to (B.10) need not always exist, one can always take a matrix $\widetilde{A}_\sigma$, $\sigma \in X$ as a solution to the minimization problem $\min_{\widetilde{A}_\sigma} ||\bar{\Gamma}\widetilde{A}_\sigma - \bar{\Gamma}_\sigma||_2$. Then $\widetilde{A}_\sigma$ can be obtained using standard numerical techniques for solving approximation problems. With this modification, the algorithm can be applied even if (B.11) fails. However, the representation returned by the modified algorithm need not be a $2N+1$-partial representation in this case.*

B.2 Proof of the partial-realization results for formal power series

The goal of the section is to present the proof of Theorem 10, 11 and 12. In Subsection B.2.1 we will introduce some notation and state some preliminary results, which will be needed for the proof of the theorems. In Subsection B.2.2 we will present the proof of Theorem 10–12. Finally, in Subsection B.2.3 we will present the proof of the technical results which are used for the proof of the Theorems 10–12.

B.2.1 Auxiliary definitions and results

To begin with, for the purposes of partial realization theory, we need to recall some basic steps of the proof of Theorem 8. To this end, we have to introduce additional notation and terminology. Let $w \in X^*$ be a word over X^* and let $S \in \mathbb{R}^p \ll X^* \gg$ be a formal power series. Define the formal power series $w \circ S \in \mathbb{R}^p \ll X^* \gg$, called the *left shift of S by w*, as follows; we require that for all $v \in X^*$ the value of $w \circ S$ at v is as follows

$$(w \circ S)(v) = S(wv) \tag{B.13}$$

i.e. the value of $w \circ S$ at v equals the value of S at wv. Notice that for any word $w \in X^*$ of the form $w = \sigma_1\sigma_2\cdots\sigma_k$, $\sigma_1, \sigma_2, \ldots, \sigma_k \in X$ and for any formal power series $T \in \mathbb{R}^p \ll X^* \gg$, the following equality holds

$$w \circ T = \sigma_k \circ (\sigma_{k-1} \circ (\cdots(\sigma_1 \circ T)\cdots))) \tag{B.14}$$

Moreover, notice that the shift operation is linear, that is, for any $T, S \in \mathbb{R}^p \ll X^* \gg$, and for any scalars $\alpha, \beta \in \mathbb{R}$, and for any word $w \in X^*$, $w \circ (\alpha S + \beta T) = \alpha(w \circ S) + \beta(w \circ T)$.

Definition 30 (Smallest shift invariant space) *Let $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ be a family of formal power series. Define the* the smallest shift invariant linear space containing all the elements of Ψ*, denoted by W_Ψ,* *as the following subspace of* $\mathbb{R}^p \ll X^* \gg$,

$$W_\Psi = \text{Span}\{w \circ S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J, w \in X^*\} \tag{B.15}$$

That is, W_Ψ *is composed of the linear combinations of all formal power series of the form* $w \circ S_j$ *for some* $j \in J$ *and* $w \in X^*$.

Remark 14 *It is easy to see that there is one-to-one correspondence between the formal power series* $w \circ S_j$ *and the column of* H_Ψ *indexed by* (w, j) *for any word* $w \in X^*$ *and index* $j \in J$. *In particular, it follows that* W_Ψ *is isomorphic to the span of columns of* H_Ψ *and hence*

$$\dim W_\Psi = \text{rank } H_\Psi$$

The statement of Theorem 8 follows from the following two auxiliary statements, proofs of which can be found in [44,42]

Lemma 2 ([44,42]) *Assume that* $\dim W_\Psi < +\infty$ *holds. Then a representation* R_Ψ *of* Ψ *is given by*

$$R_\Psi = (W_\Psi, \{A_\sigma\}_{\sigma \in X}, B, C) \tag{B.16}$$

where for each letter $\sigma \in X$, *the map* $A_\sigma : W_\Psi \to W_\Psi$ *is defined as the shift by* σ, *i.e. for each* $T \in W_\Psi$, $A_\sigma(T) = \sigma \circ T$; *the collection* $B = \{B_j \in W_\Psi \mid j \in J\}$ *is formed by the elements of* Ψ, *that is* $B_j = S_j$ *for each* $j \in J$; *the linear map* $C : W_\Psi \to \mathbb{R}^p$ *is defined as the evaluation at the empty word, i.e.* $C(T) = T(\epsilon)$ *for all* $T \in W_\Psi$.

Lemma 3 ([44,42]) *If* Ψ *is rational, then* $\dim W_\Psi < +\infty$. *More precisely, for each representation* R *of* Ψ, $\dim W_\Psi \leq \dim R$.

Using the lemmas above the proof of Theorem 8 becomes trivial. For the sake of completeness we present it below. From Remark 14 it follows that $\dim W_\Psi = \text{rank } H_\Psi$. If $\text{rank } H_\Psi < +\infty$, then Lemma 2 implies that R_Ψ is a well-defined representation of Ψ, hence Ψ is rational. Conversely, if Ψ is rational then Lemma 3 implies that $\dim W_\Psi = \text{rank } H_\Psi < +\infty$.

Remark 15 *The representation* R_Ψ *defined in (B.16) is called* free.

Below we will define various spaces which play a role analogous to W_Ψ. However, before proceeding further, some additional notation needs to be set up.

Notation 5 (Words of length at most N) *We will denote the set of all words over* X *of length at most* N *by* $X^{\leq N}$, *i.e.* $X^{\leq N} = \{w \in X^* \mid |w| \leq N\}$.

We will start with defining the space $W_{\Psi,.,N}$ of formal power series which corresponds to the subset of columns of H_Ψ indexed by indices of the form (w,j) with $j \in J$ and $w \in X^{\leq N}$. Define the vector space $W_{\Psi,.,N}$ as the subspace of W_Ψ spanned by all formal power series of the form $w \circ S_j$ with $j \in J$ and $w \in X^{\leq N}$, i.e.

$$W_{\Psi,.,N} = \text{Span}\{(w \circ S_j) \mid j \in J, w \in X^*, |w| \leq N\} \tag{B.17}$$

Notice that if J is finite, then $W_{\Psi,.,N}$ is generated by finitely many elements. The motivation for considering the space $W_{\Psi,.,N}$ is revealed by the following lemma and its corollary.

Lemma 4 *Assume that* rank $H_\Psi \leq N$*. For any formal power series $T \in W_\Psi$ and any word $w \in X^*$, there exists scalars $\alpha_{w,v} \in \mathbb{R}$, and words $v \in X^{\leq N-1}$ such that*

$$w \circ T = \sum_{v \in X^{\leq N-1}} \alpha_{w,v}(v \circ T).$$

That is, the shift of T by w is a linear combination of the shifts of T by words of the length at most $N-1$.

The proof of Lemma 4 will be presented in Subsection B.2.3.

Corollary 1 *If* rank $H_\Psi \leq N+1$*, then $W_\Psi = W_{\Psi,.,N}$.*

The proof of Corollary 1 will be presented in Subsection B.2.3. In other words, if N is big enough, then $W_{\Psi,.,.N}$ generates the whole space W_Ψ. Alternatively, stated in the language of Hankel-matrices, the columns of H_Ψ indexed by words of length at most N span the whole image of H_Ψ.

Although $W_{\Psi,.,N}$ is generated by finitely many elements, if J is finite, its elements are formal power series which contain infinite amount of data. In order to replace $W_{\Psi,.,N}$ by a space generated by finite vectors, we proceed as follows. We will define a linear space $W_{\Psi,M,N}$ which plays a similar role for $H_{\Psi,M,N}$ as W_Ψ does for H_Ψ. Similarly to W_Ψ, the elements of $W_{\Psi,M,N}$ are maps mapping words over X to vectors in $\mathbb{R}^p$. However, in contrast to W_Ψ, the elements of $W_{\Psi,M,N}$ will be defined only on the words of length at most M. Similarly to W_Ψ, the space $W_{\Psi,M,N}$ will allow us to prove our main results for the finite Hankel sub-matrices $H_{\Psi,M,N}$ in a more intuitive way.

To this end, we will introduce notation for restrictions of formal power series $\mathbb{R}^p \ll X^* \gg$ to words of length at most M.

Notation 6 (Restriction of formal power series to $X^{\leq M}$) *Denote by $\mathbb{R}^p \ll X^{\leq M} \gg$ the set of functions $T : X^{\leq M} \to \mathbb{R}^p$.*

It is clear that $\mathbb{R}^p \ll X^{\leq M} \gg$ forms a vector space with point-wise addition and point-wise multiplication by scalar.

As the next step, define the map

$$\eta_M : \mathbb{R}^p \ll X^* \gg \to \mathbb{R}^p \ll X^{\leq M} \gg$$

projecting any formal power series to its restriction to $X^{\leq M}$, i.e. for each formal power series $T \in \mathbb{R}^p \ll X^* \gg$,

$$\eta_M(T)(w) = T(w) \text{ for all } w \in X^{\leq M} \tag{B.18}$$

It is easy to see that η_M is a surjective linear map. Define the vector space $W_{\Psi,N,M}$ by

$$W_{\Psi,M,N} = \text{Span}\{\eta_M(w \circ S_j) \mid w \in X^{\leq N}, j \in J\} \tag{B.19}$$

It is easy to see that $W_{\Psi,M,N}$ is the image of $W_{\Psi,.,N}$ by η_M. The relationship between $W_{\Psi,M,N}$ and $H_{\Psi,M,N}$ can be best described as follows; the generator set of $W_{\Psi,M,N}$ is formed by the columns of $H_{\Psi,M,N}$. More precisely, recall from Subsection 4.0.2 that $\mathbb{R}^{I_M}$ denotes the vector space of all maps from I_M to $\mathbb{R}$. Define the map $\psi_M : \mathbb{R}^p \ll X^{\leq M} \gg \to \mathbb{R}^{I_M}$ by

$$(\psi_M(T))_{(v,i)} = (T(v))_i \text{ for all } (v,i) \in I_M \tag{B.20}$$

where $(T(v))_i$ stands for the ith entry of the vector $T(v) \in \mathbb{R}^p$. It is easy to see that ψ_M is a linear isomorphism. Moreover, $\text{Im} H_{\Psi,M,N} = \psi_M(W_{\Psi,M,N})$. In particular,

Corollary 2 $\dim W_{\Psi,M,N} = \text{rank } H_{\Psi,M,N}$.

Finally, we will show that if Ψ is rational and M is big enough, then the restriction of ψ_M to W_Ψ is injective. It means then that $W_{\Psi,M,N}$ and $W_{\Psi,.,N}$ are isomorphic for all N, if M is large enough.

Lemma 5 *Assume that* rank $H_\Psi = \dim W_\Psi \leq M+1$. *Then the restriction of* η_M *to* W_Ψ *is injective.*

B.2.2 Proof of the main results on partial realization theory of formal power series

Now we are ready to present the proof of Theorem 10 - 12.

PROOF. [Proof of Theorem 10]

Proof of Part 1 We will define a $2N+1$ partial representation $\hat{R}_N$ of Ψ on the linear space $W_{\Psi,N,N}$. The representation $\hat{R}_N$ will be reachable and observable and will satisfy a number of properties.

To this end, define the map

$$\eta_{N+1,N} : \mathbb{R}^p \ll X^{\leq N+1} \gg \to \mathbb{R}^p \ll X^{\leq N} \gg$$

by requiring that for all $S \in \mathbb{R}^p \ll X^{\leq N+1} \gg$ and for each word $w \in X^{\leq N}$,

$$\eta_{N+1,N}(S)(w) = S(w) \tag{B.21}$$

It is easy to see $\eta_{N+1,N}$ is a surjective linear map. Moreover, $\eta_{N+1,N}$ maps the sub-space $W_{\Psi,N+1,N}$ onto $W_{\Psi,N,N}$. Using Corollary 2, (B.2) can be rewritten as

$$\dim W_{\Psi,N,N} = \dim W_{\Psi,N+1,N} = \dim W_{\Psi,N,N+1} \tag{B.22}$$

But (B.22) implies that $W_{\Psi,N,N+1} = W_{\Psi,N,N}$ and that the restriction of $\eta_{N+1,N}$ to $W_{\Psi,N,N}$ is injective. The latter means that the inverse map

$$\eta_{N+1,N}^{-1} : W_{\Psi,N,N} \to W_{\Psi,N+1,N} \tag{B.23}$$

exists.

Define now the representation $\hat{R}_N$ as follows.

$$\hat{R}_N = (W_{\Psi,N,N+1}, \{\hat{A}_\sigma\}_{\sigma \in X}, \hat{B}, \hat{C}) \tag{B.24}$$

where

- For each letter $\sigma \in X$, the linear map

  $$\hat{A}_\sigma : W_{\Psi,N,N} = W_{\Psi,N,N+1} \to W_{\Psi,N,N+1}$$

  is defined as follows. Consider the map $T_\sigma : W_{\Psi,N+1,N} \to \mathbb{R}^p \ll X^{\leq N} \gg$ defined by $T_\sigma(Z)(v) = Z(\sigma v)$, for all $v \in X^{\leq N}$ and $Z \in W_{\Psi,N+1,N}$. It is easy to see that T_σ is a linear map. In addition, if $Z = \eta_{N+1}(w \circ S_j)$ for some $w \in X^{\leq N}$, $j \in J$, then $T_\sigma(Z) = \eta_N(w\sigma \circ S_j)$. Indeed, for any $v \in X^{\leq N}$, $T_\sigma(Z)(v) = w \circ S_j(\sigma v) = S_j(w\sigma v) = w\sigma \circ S_j(v)$. Hence, because of linearity of T_σ, the range of T_σ is $W_{\Psi,N,N+1}$ and thus T_σ can be viewed as a map $T_\sigma : W_{\Psi,N+1,N} \to W_{\Psi,N,N+1}$.

  Then for all $S \in W_{\Psi,N,N} = W_{\Psi,N,N+1}$, define

  $$\hat{A}_\sigma(S) = T_\sigma(\eta_{N+1,N}^{-1}(S))$$

  It is easy to see that $\hat{A}_\sigma$ is a well-defined linear map and it has the property that it commutes with the shift by σ of a formal power series in $W_{\Psi,.,N}$, i.e. for all $S \in W_{\Psi,.,N}$,

  $$\hat{A}_\sigma(\eta_N(S)) = \eta_N(\sigma \circ S) \tag{B.25}$$

  Here, we used the notation of (B.17). Indeed, notice that $\eta_{N+1,N}^{-1}(\eta_N(S)) = \eta_{N+1}(S)$. In addition, from the discussion above it follows that $T_\sigma(\eta_{N+1}(S)) = \eta_N(\sigma \circ S)$. Combining this with the definition of A_σ yields (B.25).

In addition, for all $S \in W_{\Psi,N,N}$, and for all $v \in X^{\leq N-1}$ and $\sigma \in X$,

$$\{\hat{A}_\sigma(S)\}(v) = \{T_\sigma(\eta^{-1}_{N+1,N}(S))\}(v) = \{\eta^{-1}_{N+1,N}(S)\}(\sigma v) = S(\sigma v) \quad \text{(B.26)}$$

- The family $\hat{B} = \{\hat{B}_j \in W_{\Psi,N,N} \mid j \in J\}$ is defined as

$$\hat{B}_j = \eta_N(S_j) \text{ for all } j \in J \quad \text{(B.27)}$$

  i.e. $\hat{B}_j$ is just the restriction of S_j to $X^{\leq N}$.
- The map $\hat{C}$ is defined as

$$\hat{C} : W_{\Psi,N,N+1} \ni S \mapsto S(\epsilon) \in \mathbb{R}^p \quad \text{(B.28)}$$

  i.e. $\hat{C}$ is just the evaluation of the elements of $W_{\Psi,N+1,N}$ at the empty word ϵ.

It is easy to see that $\hat{R}_N$ is a well-defined rational representation. Next, we will show that $\hat{R}_N$ is an $2N+1$-partial representation of Ψ. To this end, by repeated application of (B.25) and using the equality $W_{\Psi,N,N} = W_{\Psi,N,N+1}$ we get that for any word $w \in X^{\leq N+1}$ it holds that

$$\hat{A}_w \hat{B}_j = \hat{A}_w(\eta_N(S_j)) = \eta_N(w \circ S_j) \quad \text{(B.29)}$$

Using (B.29) and repeatedly applying (B.26) yields that for each $v \in X^{\leq N}$,

$$\hat{C}\hat{A}_v\hat{A}_w\hat{B}_j = \{\hat{A}_v \eta_N(w \circ S_j)\}(\epsilon) = \{w \circ S_j\}(v) = S_j(wv) \quad \text{(B.30)}$$

By noticing that $\hat{A}_v\hat{A}_w = \hat{A}_{wv}$ and that any word $\hat{w}$ in $X^{\leq 2N+1}$ can be represented as the concatenation $\hat{w} = wv$ of a word $w \in X^{\leq N+1}$ with a word $v \in X^{\leq N}$, we get that for all $\hat{w} \in X^{\leq 2N+1}$,

$$CA_{\hat{w}}\hat{B}_j = S_j(\hat{w})$$

That is, $\hat{R}_N$ is an $2N+1$-partial representation of Ψ. From (B.29) it is easy to deduce that $\hat{R}_N$ is reachable. Observability of $\hat{R}_N$ can be derived as follows. From (B.26) it follows that for any $T \in W_{\Psi,N,N+1} = W_{\Psi,N,N}$, $\hat{C}\hat{A}_wT = T(w)$ for all words $w \in X^{\leq N}$. Hence, if $\hat{C}\hat{A}_w(T) = 0$ for all $w \in X^*$, then for all $w \in X^{\leq N}$, $T(w) = 0$, i.e. $T = 0$. That is, $O_{\hat{R}_N} = \{0\}$.

Define now the representation R_N as the isomorphic copy of $\hat{R}_N$ with the isomorphism ψ_N defined in (B.20). That is, using the notation of the theorem, define the parameters A_σ, $\sigma \in X$, C and $\{B_j \mid j \in J\}$ of R_N as follows.

$$A_\sigma = \psi_N \hat{A}_\sigma \psi_N^{-1} \text{ for all } \sigma \in X,\ C = \hat{C}\psi_N^{-1},\ B_j = \psi_N(\hat{B}_j) \text{ for all } j \in J. \quad \text{(B.31)}$$

Since ψ_N defines a representation isomorphism $\psi_N : \hat{R}_N \to R_N$, it follows that R_N is a $2N+1$-rational representation of Ψ as well, and it is reachable and observable. In addition, combining (B.31) with (B.25,B.27,B.28) yields (B.4,B.5,B.6).

Proof of Part 2
Recall that rank $H_{\Psi,M,K} = \dim W_{\Psi,M,K}$ for all $K, M \in \{N, N+1\}$ and $\dim W_{\Psi} = \text{rank } H_{\Psi}$. It is easy to see that

$$\dim W_{\Psi,N,N} \leq \dim W_{\Psi,K,M} \leq \dim W_{\Psi} \text{ for all } K, M \in \{N, N+1\} \quad \text{(B.32)}$$

Hence, (B.7) implies that

$$\dim W_{\Psi,N,N} = \dim W_{\Psi,M,K} = \dim W_{\Psi} \text{ for all } M, K \in \{N, N+1\}$$

Hence, (B.2) holds and by Part 1 the representation R_N is well-defined and it is a $2N+1$-partial representation of Ψ. Recall that (B.2) implies that $W_{\Psi,N,N+1} = W_{\Psi,N,N}$.

It is left to show that R_N is a representation of Ψ. To this end, recall from (B.18) the definition of the map η_N and recall from (B.17) the definition of the set $W_{\Psi,.,N}$. Since $\eta_N(W_{\Psi,.,N}) = W_{\Psi,N,N}$, we get that

$$\dim W_{\Psi} \geq \dim W_{\Psi,.,N} \geq \dim W_{\Psi,N,N} = \dim W_{\Psi},$$

hence $W_{\Psi,.,N} = W_{\Psi}$. Notice that η_N maps $W_{\Psi,.,K}$ onto $W_{\Psi,N,K}$ for $K \in \{N, N+1\}$. Hence, the map η_N maps W_{Ψ} onto $W_{\Psi,N,N} = W_{\Psi,N,N+1}$. That is, the restriction η_N to W_{Ψ} yields a linear isomorphism $\eta_N|_{W_{\Psi}} : W_{\Psi} \to W_{\Psi,N,N+1}$. Recall from Lemma 2 the definition of the free representation R_{Ψ} of Ψ, defined on the space W_{Ψ}. Using (B.25),(B.27) and (B.28), it is easy to see that the restriction of η_N to W_{Ψ} in fact yields a representation isomorphism $\eta_N|_{W_{\Psi}} : R_{\Psi} \to \hat{R}_N$, where $\hat{R}_N$ is as defined in (B.24). Since, R_{Ψ} is a minimal representation of Ψ, then so is its isomorphic copy $\hat{R}_N$. Since R_N is merely an isomorphic copy of $\hat{R}_N$, the same conclusion holds for R_N.

Proof of Part 3
Recall from (B.17) and (B.19) the definition of the spaces $W_{\Psi,.,N}$ and $W_{\Psi,M,N}$ for $M \in \mathbb{N}$, $M > 0$. From Lemma 3 it follows that if $\dim R \leq N+1$, then $\dim W_{\Psi} = \text{rank } H_{\Psi} \leq \dim R \leq N+1$. Then by Corollary 1, $W_{\Psi,.,K} = W_{\Psi}$ for $K = N, N+1$. Moreover, notice that the image of $W_{\Psi,.,N}$ by η_M equals $W_{\Psi,M,N}$. Applying the above observations for $M, K \in \{N, N+1\}$ and using Lemma 5 we get that η_N and η_{N+1} are injective and hence

$$\dim W_{\Psi} = \dim W_{\Psi,.,K} = \dim W_{\Psi,M,K} \text{ where } K, M = N, N+1 \quad \text{(B.33)}$$

Combining (B.33) with Corollary 2 and Remark 14 we get that (B.7) holds. Combining this and Part 2 we get that the statement of Part 3 holds.

Next, we will present the proof of Theorem 11.

PROOF. [Proof of Theorem 11] We will prove the statements of the theorem one by one. However, before proceeding to the actual proof, we need to introduce some notation. Let R be p-J rational representation, i.e. R is of the form (A.1). Define the family of formal power series Ψ_R *associated with the representation* R as follows. The family Ψ_R is indexed by elements of J, i.e. it can be written as $\Psi_R = \{S_j^R \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$. For each index $j \in J$, and for each word $w \in X^*$, the value of the formal power series S_j^R at w is defined as $S_j^R(w) = CA_wB_j$. It is easy to see that the rational representation R is a representation of Ψ_R, i.e. Ψ_R is rational. In fact, R is a representation of the family of formal power series Ψ, if and only if $\Psi_R = \Psi$, i.e. $S_j = S_j^R$ for all $j \in J$. Furthermore, R is a $2N+1$-partial representation of Ψ, if and only if $S_j^R(w) = S_j(w)$ for all $j \in J$ and $w \in X^{\leq 2N+1}$. In addition, if R is a $2N+1$-partial representation of Ψ, then

$$H_{\Psi,K,L} = H_{\Psi_R,K,L} \text{ for all } K + L \leq 2N+1 \tag{B.34}$$

In particular, $H_{\Psi,N,N} = H_{\Psi_R,N,N}$. Now we are ready to proceed to the actual proof of the theorem.

Proof of Part 1

Assume that R is a $2N+1$-partial representation of Ψ. From the discussion above it follows that $\text{rank } H_{\Psi,N,N} = \text{rank } H_{\Psi_R,N,N}$. It is easy to see that $\text{rank } H_{\Psi_R,N,N} \leq \text{rank } H_{\Psi_R}$. Moreover, from Theorem 9 it follows that $\text{rank } H_{\Psi_R} \leq \dim R$. That is, the rank of $H_{\Psi,N,N}$ is not greater than the dimension of the representation R. But from Theorem 10 it follows the the dimension of the representation R_N is precisely $\text{rank } H_{\Psi,N,N}$ and that R_N is an $2N+1$-partial representation of Ψ. Hence, we get that $\dim R_N \leq \dim R$. In other words, R_N is a minimal $2N+1$ representation of Ψ.

Proof of Part 2

We have just shown that the rational representation R_N of Theorem 10 is a minimal $2N+1$-partial representation of Ψ. If R is another minimal $2N+1$-partial representation of Ψ, then the dimension of R must be equal to that of R_N, which in turn equals $\text{rank } H_{\Psi,N,N}$. Hence, the second part of the statement is proved. Assume now that R is a minimal $2N+1$-partial representation of Ψ. Then it means that $S_j^R(w) = S_j(w)$ for each $j \in J$ and $w \in X^{\leq 2N+1}$, where S_j^R denotes the element of Ψ_R indexed by the index $j \in J$. If R is not reachable and observable, then it is not a minimal representation of Ψ_R. From Theorem 9 it follows then that there exists a reachable and observable representation R_m of Ψ_R such that $\dim R_m < \dim R$. But if $R_m = (\mathcal{X}^m, \{A_\sigma^m\}_{\sigma \in X}, B^m, C^m)$ is a representation of Ψ_R, then it is a $2N+1$-partial representation of Ψ, since, $S_j(w) = S_j^R(w) = C^m A_w^m B^m$ for all $j \in J$ and $w \in X^{\leq 2N+1}$. Hence, R is not a minimal $2N+1$-partial representation of Ψ, which is a contradiction.

Proof of Part 3

Suppose R and $\hat{R}$ are two minimal $2N+1$-partial representations of Ψ. Recall the definition of the families of formal power series Ψ_R and $\Psi_{\hat{R}}$ associated with R. Since R and $\hat{R}$ have to be reachable and observable by Part 2 of the current theorem, by Theorem 9 the representation R (respectively $\hat{R}$) is a minimal representation of Ψ_R (respectively $\Psi_{\hat{R}}$). We will show that Ψ_R and $\Psi_{\hat{R}}$ are equal, i.e. $S^R_j = S^{\hat{R}}_j$ for all $j \in J$. This in turn implies that R and $\hat{R}$ are both minimal representations of the same family of formal power series, and hence by Theorem 9 they are isomorphic.

The proof that Ψ_R and $\Psi_{\hat{R}}$ are identical proceeds as follows. Recall that R and $\hat{R}$ are both of dimension rank $H_{\Psi,N,N}$, and recall (B.34). Hence, it follows that rank $H_{\Psi,N,N} = \text{rank } H_{\Psi_R,N,N} = H_{\Psi_R} = \dim R$ and rank $H_{\Psi,N,N} = \text{rank } H_{\Psi_{\hat{R}},N,N} = \text{rank } H_{\Psi_{\hat{R}}} = \dim \hat{R}$. It also follows that Ψ_R and $\Psi_{\hat{R}}$ satisfy (B.7) with Ψ replaced by Ψ_R and $\Psi_{\hat{R}}$ respectively. Denote by R^1_N and R^2_N the minimal $2N+1$-partial representations of Ψ_R and respectively $\Psi_{\hat{R}}$, obtained by applying Part 1 of Theorem 10 to Ψ_R and $\Psi_{\hat{R}}$ respectively. By Part 2 of Theorem 10 the representations R^1_N and R^2_N are minimal representations of Ψ_R and $\Psi_{\hat{R}}$ respectively. However the construction of R^1_N and R^2_N depends only on $H_{\Psi_R,N,N+1}$ and $H_{\Psi_{\hat{R}},N,N+1}$. The latter two matrices both coincide with $H_{\Psi,N,N+1}$. Hence, R^1_N and R^2_N are both equal to the partial representation R_N obtained by applying Part 1 of Theorem 10 to Ψ. That is, rational representation R_N, obtained by applying Part 1 of Theorem 10 to Ψ, is a minimal representation of both Ψ_R and $\Psi_{\hat{R}}$. This implies that Ψ_R and $\Psi_{\hat{R}}$ coincide.

We conclude with the proof of Theorem 12.

PROOF. [Proof of Theorem 12]

Proof of Part 1
We have to show that Algorithm 3 returns a representation $\widetilde{R}_N$, and $\widetilde{R}_N$ is isomorphic to R_N. By Theorem 10, if (B.11) holds then R_N is an $2N+1$-representation of Ψ. From Theorem 11 it then follows that R_N is a minimal $2N+1$-partial representation. Hence, if $\widetilde{R}_N$ is an isomorphic copy of R_N, then we get that $\widetilde{R}_N$ is a minimal $2N+1$-partial representation of Ψ.

Recall from (B.24) in the proof of Theorem 10 the definition of the representation $\hat{R}_N$. Using the remark in Subsection 4.0.2, we will interpret the matrix $O \in \mathbb{R}^{I_{N+1} \times r}$ as a linear map $O : \mathbb{R}^r \to \mathbb{R}^{I_{N+1}}$ defined by $(Ox)_i = \sum_{k=1}^r O_{i,k} x_k$ for all $x \in \mathbb{R}^r$. Using this interpretation, we define the map

$$\hat{\xi} = \eta_{N,N+1} \circ \psi^{-1}_{N+1} \circ O : \mathbb{R}^r \to W_{\Psi,N,N} = W_{\Psi,N,N+1} \tag{B.35}$$

Since ψ_N defines a representation isomorphism $\psi_N : \hat{R}_N \to R_N$ it is enough to show that $\hat{\xi}$ defines a representation isomorphism $\hat{\xi} : \widetilde{R}_N \to \hat{R}_N$ and that

Algorithm 3 returns a rational representation if (B.11) holds. Indeed, in this case $\xi = \psi_N \circ \hat{\xi}$ defines a representation isomorphism $\xi : \widetilde{R}_N \to R_N$.

It is clear that $\hat{\xi}$ is well defined. Indeed, $\mathrm{Im}O = \mathrm{Im}H_{\Psi,N+1,N}$ by definition of matrix factorization. Moreover, $O : \mathbb{R}^r \to \mathbb{R}^{I_{N+1}}$ is an injective linear map and

$$\psi_{N+1} : W_{\Psi,N+1,N} \to \mathrm{Im}H_{\Psi,N+1,N}$$

is a linear isomorphism. Furthermore, since

$$\mathrm{rank}\ H_{\Psi,N+1,N} = \dim W_{\Psi,N+1,N} = \mathrm{rank}\ H_{\Psi,N,N} = \dim W_{\Psi,N,N}$$

we get that $\eta_{N+1,N} : W_{\Psi,N+1,N} \to W_{\Psi,N,N}$ is a linear isomorphism. Thus, $\hat{\xi} = \eta_{N+1,N} \circ \psi^{-1}_{N+1} \circ O$ is a well defined linear isomorphism. It is left to show that $\hat{\xi}$ is a representation morphism from $\widetilde{R}_N$ to $\hat{R}_N$.

It is easy to see that for all $w \in X^{\leq N}$, and $x \in \mathbb{R}^r$.

$$\hat{\xi}(x)(w) = \Big[(Ox)_{(w,1)},\ (Ox)_{(w,2)},\ \cdots\ (Ox)_{(w,p)}\Big]^T \tag{B.36}$$

Recall from Subsection 4.0.2 that $(Ox)_{(w,i)}$ stands for the value (entry) of $Ox \in \mathbb{R}^{I_{N+1}}$ corresponding to the index $(w, i) \in I_{N+1}$. In particular, for any $(v, j) \in J_N$,

$$\hat{\xi}(R_{.,(v,j)}) = \eta_N(v \circ S_j) \tag{B.37}$$

where $R_{.,(v,j)}$ stands for the column of R indexed by (v, j). Indeed, using (B.36) we get that for all $w \in X^{\leq N}$,

$$\hat{\xi}(R_{.,(v,j)})(w) = \Big[(OR)_{(w,1),(v,j)},\ (OR)_{(w,2),(v,j)},\ \cdots\ (OR)_{(w,p),(v,j)}\Big]^T \tag{B.38}$$

But $H_{\Psi,N+1,N} = OR$, and hence the right-hand side of (B.38) equals the vector formed by the entries $(H_{\Psi,N+1,N})_{(w,i),(v,j)}$ for $i = 1, 2 \ldots, p$. But the entry $(H_{\Psi,N+1,N})_{(w,i),(v,j)}$ equals the ith entry of $S_j(vw)$, hence

$$\hat{\xi}(R_{.,(v,j)})(w) = S_j(vw) = v \circ S_j(w)$$

From (B.36) it follows that

$$\hat{C}(\hat{\xi}(x)) = \hat{\xi}(x)(\epsilon) = \widetilde{C}x \tag{B.39}$$

From (B.37), it follows that

$$\hat{\xi}(\widetilde{B}_j) = \eta_N(S_j) = \hat{B}_j \tag{B.40}$$

It is left to show that a unique solution to equation (B.10) exists and

$$\hat{A}_\sigma \hat{\xi} = \hat{\xi} \widetilde{A}_\sigma \tag{B.41}$$

for all $\sigma \in X$. First, notice that $\bar{\Gamma} R = H_{\Psi,N,N}$. Thus, rank $\bar{\Gamma} R =$ rank $H_{\Psi,N,N} = r$, i.e. rank $\bar{\Gamma} = r$. Hence, if a solution solution (B.10) exists, then this solution is unique. Therefore, if we show that $\hat{\xi}^{-1}\hat{A}_\sigma\hat{\xi}$ is a solution to (B.10) then (B.41) automatically holds. Notice that $\hat{\xi}^{-1}\hat{A}_\sigma\hat{\xi}$ is a linear map from $\mathbb{R}^r$ to $\mathbb{R}^r$ and hence it can be identified with its matrix representation. By identifying $\hat{\xi}^{-1}\hat{A}_\sigma\hat{\xi}$ with its matrix, and using (B.37) and (B.25), it follows that for any $(v,j) \in J_N$ and $(w,i) \in I_N$,

$$\begin{aligned}
&(\bar{\Gamma}\hat{\xi}^{-1}\hat{A}_\sigma\hat{\xi}R)_{(w,i),(v,j)} = \bar{\Gamma}_{(w,i),.}(\hat{\xi}^{-1}\hat{A}_\sigma\hat{\xi}R)_{.,(v,j)} = \\
&\bar{\Gamma}_{(w,i),.}(\hat{\xi}^{-1}\hat{A}_\sigma\hat{\xi}R_{.,(v,j)}) = \bar{\Gamma}_{(w,i),.}\hat{\xi}^{-1}\hat{A}_\sigma\eta_N(v \circ S_j) = \\
&\bar{\Gamma}_{(w,i),.}\hat{\xi}^{-1}\eta_N(v\sigma \circ S_j)
\end{aligned} \tag{B.42}$$

Here we used the notation of Subsection 4.0.2 to denote rows and columns of matrices. In particular, $M_{.,(v,j)}$ denotes the column of a matrix M indexed by (v,j) and $M_{(w,i),.}$ denotes the row of a matrix M indexed by (w,i). Notice that if $w \in X^{\leq N}$, then the row of $\bar{\Gamma}$ indexed by (w,i) equals the row of O indexed by (w,i). Moreover, if $x = \hat{\xi}^{-1}(\eta_N(v\sigma \circ S_j))$, then from (B.36) it follows that the row of $O(x)$ indexed by (w,i) equals the ith entry of $\eta_N(v\sigma \circ S_j)(w) = S_j(v\sigma w)$. Hence, the last expression of (B.42) can be rewritten as

$$\bar{\Gamma}_{(w,i),.}\hat{\xi}^{-1}\eta_N(v\sigma \circ S_j) = O(\hat{\xi}^{-1}\eta_N(v\sigma \circ S_j))_{(w,i)} = (S_j(v\sigma w))_i \tag{B.43}$$

where $(S_j(v\sigma w))_i$ denotes the ith entry of $S_j(v\sigma w) \in \mathbb{R}^p$. On the other hand, the row of $\bar{\Gamma}_\sigma$ indexed by (w,i) equals the row of O indexed by $(\sigma w, i)$, if $w \in X^{\leq N}$. Hence,

$$(\bar{\Gamma}_\sigma R)_{(w,i),(v,j)} = (O)_{(\sigma w,i),.}R_{.,(v,j)} = (H_{\Psi,N+1,N})_{(\sigma w,i),(v,j)} = (S_j(v\sigma w))_i \tag{B.44}$$

Combining (B.42), (B.43) and (B.44) we get that $\bar{\Gamma}\hat{\xi}^{-1}\hat{A}_\sigma\hat{\xi}R = \bar{\Gamma}_\sigma R$. Since rank $R = r$, i.e the columns of R span the whole space $\mathbb{R}^r$, the last equality implies that $\hat{\xi}^{-1}\hat{A}_\sigma\hat{\xi}$ is a solution to (B.10).

Proof of Part 2
From Theorem 10 it follows that if (B.12) holds, then (B.11) holds and R_N is a minimal representation of Ψ. By Part 1 of this theorem, if (B.11) holds, then Algorithm 3 returns a representation $\widetilde{R}_N$ and $\widetilde{R}_N$ is isomorphic to R_N. Hence $\widetilde{R}_N$ is a minimal representation too.

Proof of Part 3
Again, from Theorem 10 it follows that (B.12) holds in this case. The statement follows now from (B.12) and Part 2 of this theorem.

B.2.3 Proof of the auxiliary results

We conclude the section with presenting the proof of Lemma 4 and Lemma 5 used in the proof of Theorem 10. The proof relies on the following chain of results, which are interesting on their own right.

Lemma 6 *Let $\mathcal{X}$ be finite-dimensional vector space,* $\dim \mathcal{X} \leq N$*. Let $A_\sigma : \mathcal{X} \to \mathcal{X}$, $\sigma \in X$ be a family of linear maps. Then for each $y \in \mathcal{X}$, for each $w \in X^*$, the vector $A_w y$ is a linear combination of the vectors $A_v y$, for finitely many words $v \in X^{\leq N-1}$.*

PROOF. If $|w| < N$, then the statement of the lemma is trivially true. First we prove the lemma for $|w| = N$. Assume that $w = \sigma_1\sigma_2 \cdots \sigma_N, \sigma_1, \sigma_2, \ldots, \sigma_N \in X$. Consider the elements $A_{\sigma_1\sigma_2\cdots\sigma_i} y$, $i = 0, \ldots, N$. Since every $N+1$ elements of $\mathcal{X}$ are linearly dependent, we get that there exist $i = 0, \ldots, N$ such that $A_{\sigma_1\cdots\sigma_i} y$ is a linear combination of $A_{\sigma_1\cdots\sigma_j} y$, $j = 0, 2, \ldots, i-1$. Then we get that $A_w y = A_{\sigma_{i+1}\cdots\sigma_N}(A_{\sigma_1\cdots\sigma_i} y)$ is a linear combination of vectors of the form $A_{\sigma_1\cdots\sigma_j\sigma_{i+1}\cdots\sigma_N} y$, $j = 0, \ldots, i-1$. Notice that each word $\sigma_1 \cdots \sigma_j \sigma_{i+1} \cdots \sigma_N$ is of length at most $N-1$. To prove the lemma for arbitrary $|w| \geq N$ we proceed by induction on $|w| - N$. The case of $|w| = N$ we proved above. Assume that the statement of the lemma holds for $|w| \leq n + N$. Assume now that w is of the form $w = s\sigma$, where $s \in X^*$, $|s| \leq n + N$ and $\sigma \in X$ is the last letter of w. From the induction hypothesis it follows that $A_s y$ is a linear combination of vectors of the form $A_v y$, $v \in X^{\leq N-1}$. But then $A_w y = A_\sigma(A_s y)$ is a linear combination of vectors of the form $A_\sigma A_v y = A_{v\sigma} y$. For each $v \in X^{\leq N-1}$, either $v\sigma \in X^{\leq N-1}$ or $|v\sigma| = N$. In the latter case, by induction hypothesis $A_{v\sigma} y$ is again a linear combination of vectors of the form $A_{\hat{v}} y$, $\hat{v} \in X^{\leq N-1}$. Altogether, we get that $A_w y$ is a linear combination of the vectors $A_v y$, $v \in X^{\leq N-1}$.

The lemma above yields the following characterization of the reachability and observability subspaces (see Definition 25 and 26) of a rational representation.

Corollary 3 *Consider a $p - J$ representation $R = (\mathcal{X}, \{A_\sigma\}_{\sigma \in X}, B, C)$. Assume that* $\dim R \leq N$*. With the notation of (A.4–A.5), the following holds.*

$$O_R = \bigcap_{v \in X^{\leq N-1}} \ker CA_v$$
$$W_R = \operatorname{Span}\{A_v B_j \mid j \in J, v \in X^{\leq N-1}\}$$

If $\mathcal{X} = \mathbb{R}^n$ and J is finite, then the above corollary states that the observability subspace O_R is the kernel of a finite matrix, and the reachability subspace W_R is the image of a finite matrix.

PROOF. [Proof of Corollary 3] It is clear that $O_R \subseteq \bigcap_{v \in X^{\leq N-1}} \ker CA_v$. We will show that the reverse inclusion holds as well. Assume that $x \in \mathcal{X}$ is such that $CA_v x = 0$ for all $v \in X^{\leq N-1}$. By Lemma 6, for each $w \in X^*$, $A_w x$ is a linear combination of $A_v x$ for $v \in X^{\leq N-1}$. Hence, $CA_w x$ is a linear combination of $CA_v x$, $v \in X^{\leq N-1}$ and the latter implies that $CA_w x = 0$. That is, $\bigcap_{v \in X^{\leq N-1}} \ker CA_v \subseteq O_R$.

Similarly, it is clear from the definition that $\mathrm{Span}\{A_v B_j \mid v \in X^{\leq N-1}, j \in J\} \subseteq W_R$. The reverse inclusion follows from Lemma 6, according to which for each word $w \in X^*$, $A_w B_j$ is a linear combination of the vectors $A_v B_j$, $v \in X^{\leq N-1}$. Hence, we get that $W_R \subseteq \mathrm{Span}\{A_v B_j \mid v \in X^{\leq N-1}, j \in J\}$.

Finally, we will present the proof of Lemma 4 and Lemma 5.

PROOF. [Proof of Lemma 4] We will use the fact that $\dim W_\Psi = \mathrm{rank}\ H_\Psi \leq N+1$. Consider the free representation $R_\Psi = (W_\Psi, \{A_\sigma\}_{\sigma \in X}, B, C)$ of Ψ defined in Theorem 8. Notice that for any word X^*, $w \circ S_j = A_w B_j$. Apply Lemma 6 to W_Ψ, $A_\sigma : W_\Psi \to W_\Psi$, $\sigma \in X$ and $y = B_j$. Then we get that for any word $w \in X^*$, the formal power series $A_w B_j = w \circ S_j$ is a linear combination of the formal power series $A_v B_j = v \circ S_j$ with $v \in X^{\leq N}$.

PROOF. [Proof of Lemma 5] It is easy to see that η_M is a surjective linear map. Consider the free realization $R_\Psi = (W_\Psi, \{A_\sigma\}_{\sigma \in X}, B, C)$ of Ψ. From Theorem 9 we know that R_Ψ is minimal and therefore it is reachable and observable, i.e. $O_{R_\Psi} = \{0\}$. From Corollary 3 we also know that if $\dim R_\Psi = \dim W_\Psi = \mathrm{rank}\ H_\Psi \leq M+1$, then $O_{R_\Psi} = \bigcap_{v \in X^{\leq M}} \ker CA_v$. Consider the kernel of η_M. For any formal power series $S \in W_\Psi$, $\eta_M(S) = 0$ if and only if $S(w) = 0$ for all $w \in X^{\leq M}$. Hence, $CA_w S = S(w) = 0$ for each $w \in X^{\leq M}$, i.e. $S \in O_{R_\Psi} = \{0\}$. Thus, η_M is injective.