A Behavioral Approach to Linear Exact Modeling

A. C. Antoulas, *Fellow, IEEE* and J. C. Willems, *Fellow, IEEE*

***Abstract*—The *behavioral approach* to system theory provides a *parameter-free* framework for the study of the general problem of linear exact modeling and recursive modeling. The main contribution of this paper is the solution of the (continuous-time) polynomial-exponential time series modeling problem. Both recursive and nonrecursive solutions are provided and classified according to properties like complexity and controllability. It is shown in particular, that recursive modeling corresponds to updating by means of a cascade interconnection of systems. As a special case the solution of several other problems, like rational interpolation, realization, modeling of arbitrary discrete-time time series, is obtained.**

I. PROLEGOMENA

A. Overview

THE problem of identifying a model of a dynamical system from observed responses goes back a very long way and has given rise to the field of system identification. For example, the derivation of Newton's inverse-square gravitational law as suggested by Kepler's laws, can be considered as a question of system identification. In recent decades, the standard approach in this area has been to interpret the measurements as the input and the output variables of a stochastic system. The output measurements are then explained by the model through the inputs and the stochastic noise. See Ljung [22] for an authoritative account of the theory and the algorithms resulting from this mode of thinking.

Recently, we have put forward the so-called *behavioral framework* for the study of dynamical systems. It is deterministic at the outset (but stochastic generalizations are being pursued) and lends itself very well to the problem of obtaining dynamical models from observed responses. See Willems [29] for an exposition of this theory.

In the classical framework, see e.g., Kalman, Falb, and Arbib [21, chapter 1], a dynamical system is viewed as a mapping which transforms inputs $\mathbf{u}$ into outputs $\mathbf{y}$. Two basic considerations express the need for a framework at a more fundamental level. *First*, in many cases (think, for example, of electrical circuits) the distinction between inputs and outputs is not *a priori* clear; instead, it should follow as a consequence of the modeling. *Second*, it is desirable to be able to treat the different representations of a given system (for example: input–output and state-space representations) in a unified way.

In the behavioral setting, the basic variables considered are the *external or manifest* variables $\mathbf{w}$, which consist of $\mathbf{u}$ and $\mathbf{y}$, without distinguishing between them. The *collection of trajectories* describing the evolution of $\mathbf{w}$ over time defines a dynamical system. It turns out that this definition provides the right level of abstraction, necessary for accommodating the two considerations layed out above. This establishes the foundations of a *parameter-free* theory of dynamical systems, the advantages of representation-independent results—or vice versa, the disadvantages of representation-dependent results—being well recognized. The resulting central object is the *most powerful unfalsified model* (MPUM) derived from the data, which, again, is a space of trajectories. Subsequently, inputs and outputs can be introduced and the corresponding I/O operator recovered.

The idea of using this approach in order to obtain dynamical models from a set of observed time series has been pursued since the very first publications related to this approach. Both the situation of exact (noiseless) modeling, cf. Willems [26], and of approximate modeling, cf. Willems [27], have been treated; the results of this latter paper were generalized in the monograph by Heij [18].

In the present paper we will exclusively treat exact modeling. The *data* (observed responses) are continuous-time polynomial-exponential time series, and the *model class* consists of linear, continuous-time, time invariant, finite dimensional systems. The novel features are i) the specialization of the general behavioral framework to the case of polynomial-exponential data, ii) the treatment of the modeling question for *any a priori* imposed input–output structure, iii) the fact that the *controllability* properties of the resulting model are worked out, thus making contact with transfer function fitting, and iv) the *recursivity* of the algorithms and their system theoretic interpretation.

Two problems which have been studied extensively, and can be cast into the framework of exact modeling, are

Manuscript received November 20, 1991; revised December 1, 1992. Paper recommended by Past Associated Editor, J. Hammer. The research of the first author was supported in part by the NSF under Grant ECS-9017618 and by the State of Texas through the Advanced Technology Program under Grant #003604-041. The research of the second author was supported in part under Grant SC1*-CT92-0779 on System Identification of the Science Program of the Commission of the European Community.

A. C. Antoulas is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251 and with the Mathematical System Theory, ETH Zurich, CH-8092 Zurich, Switzerland.

J. C. Willems is with the Mathematics Institute, University of Groningen, 9700 AV Groningen, The Netherlands.

IEEE Log Number 9213188.

0018-9286/93$03.00

realization[1] and *rational interpolation*. Roughly speaking, the data in the former problem are impulse response measurements, while in the latter, they are frequency response measurements. As shown subsequently, the realization problem can be interpreted as a special case of the interpolation problem, which in turn, is a special case of the general polynomial-exponential time series modeling problem studied here (see Section II). At this stage, we would like to draw the reader's attention to Antoulas [1]; this paper studies the recursive realization problem, and in many respects, its results anticipate the recursiveness results presented below. The main result of the above paper, for example, which is shown pictorially in Fig. 3, remains valid in the present context: *recursive update corresponds to attaching an appropriately defined component to a cascade interconnection of systems* (for details see Section VIII–X and Remark 6.9).

B. The Data

We will denote by $\mathbf{k}$ the field of real numbers $\mathbf{R}$ or that of complex numbers $\mathbf{C}$. The following finite collection of vector-valued continuous-time polynomial-exponential time series, with domain $\mathbf{T} = \mathbf{R}$ and range $\mathbf{k}^q$, constitutes our *data set*:

$$\mathbf{D} := \{w_i := \mathbf{p}_i \exp_{\lambda_i}, \; i \in \underline{n}\};^2 \tag{1.1}$$

$\mathbf{p}_i$ is a vector-valued polynomial function:

$$\mathbf{p}_i: \mathbf{R} \to \mathbf{k}^q \qquad \text{where } t \mapsto \mathbf{p}_i(t) := \sum_{j=1}^{\kappa_i} p_{i,\kappa_i - j} \frac{t^{j-1}}{(j-1)!}, \quad p_{i,0} \neq 0, \; p_{i,\kappa_i - j} \in \mathbf{k}^q; \tag{1.2}$$

the function $\exp_\lambda$ is defined as follows:

$$\exp_\lambda: \mathbf{R} \to \mathbf{k} \qquad \text{where } t \mapsto \exp_\lambda(t) := e^{\lambda t} \text{ with } \lambda \in \mathbf{k}.^2 \tag{1.3}$$

The constant λ_i is called the *frequency* of the time series $\mathbf{w}_i$. For subsequent use we define the integers:

$$N := \sum_{i=1}^{n} \kappa_i \qquad \text{where } \kappa_i := \deg \mathbf{p}_i + 1. \tag{1.4}$$

C. The Main Problem (Rough Statement)

Given the above data set $\mathbf{D}$, find *all* linear, continuous-time, time invariant, finite dimensional systems which *explain*, i.e., could have generated, the data $\mathbf{w}_i \in \mathbf{D}$, $i \in \underline{n}$. In particular, modeling algorithms which are *recursive* in the number of time series n are to be devised. The important parameters to keep track of in these considerations are:

model complexity and *model controllability*.

Other issues of interest are

imposed input–output model structure and *model smoothness*.

The precise formulation of the *main problem* is given in Section III-F, following a brief introduction to the behavioral framework.

D. Overview of Contents

1) The Basic Idea

Consider the problem defined in Sections I-B and I-C with $q = 2$. In this case, each measurement $w_i \in \mathbf{D}$ can be written as $\mathbf{w}_i = \begin{pmatrix} \mathbf{u}_i \\ \mathbf{y}_i \end{pmatrix}$, and can be interpreted as a measurement on a single-input single-output system. The problem of finding *all* linear, time invariant systems which are compatible with the given data $\mathbf{D}$, reduces to finding all polynomials $\mathbf{p}, \mathbf{q} \in \mathbf{k}[s]$ such that

$$\left(\mathbf{p}\left(\frac{d}{dt}\right) \quad -q\left(\frac{d}{dt}\right)\right)\mathbf{w}_i = 0, \qquad i \in \underline{n}.$$

The basic concept behind the solution of this problem is what we shall call the *generating system*, which in the case $q = 2$ is a 2×2 nonsingular polynomial matrix

$$\Theta^* := \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix} \in \mathbf{k}^{2\times 2}[s], \qquad \det \Theta^* \neq 0,$$

computed directly from the data $\mathbf{D}$. The main result is that $\mathbf{p}, \mathbf{q}$ is a solution pair if, and only if, there exists a pair of polynomial $\mathbf{a}, \mathbf{b}$ such that

$$(\mathbf{p} \quad -\mathbf{q}) = (\mathbf{a} \quad -\mathbf{b})\Theta^* = (\mathbf{a}\theta_{11} - \mathbf{b}\theta_{21} \quad \mathbf{a}\theta_{12} - \mathbf{b}\theta_{22}).$$

If $\mathbf{p}, \mathbf{q}$ are co-prime, the solution can be equivalently described in terms of the transfer function

$$\frac{\mathbf{p}}{\mathbf{q}} = \frac{\mathbf{a}\theta_{11} - \mathbf{b}\theta_{21}}{\mathbf{a}\theta_{12} - \mathbf{b}\theta_{22}}$$

which is a *linear fraction*. Thus while the generating system Θ^* depends on $\mathbf{D}$, the polynomials $\mathbf{a}, \mathbf{b}$ are the free parameters. The above solution has an interpretation in terms of the feedback interconnection of the systems Σ defined by Θ^*, and $\bar{\Sigma}$ defined by $\Gamma := (\mathbf{a} - \mathbf{b})$, as shown in Fig. 2 and explained in Section X.

Through the years, solutions to various special cases of the general polynomial-exponential modeling problem have been worked out in what amounts to a generating system approach. For example, more than three quarters of a century ago, one version of the rational interpolation problem, the so-called (scalar) *Nevanlinna–Pick* or constrained interpolation problem, was solved using this approach. Actually, the generating system was constructed recursively. For details on this problem cast in the generating system framework see, e.g., Antoulas and Anderson [3], or Ball, Gohberg, and Rodman [12].

More recently, the general unconstrainted rational interpolation problem in all its matrix and tangential versions, has also been solved in the generating system framework, with the complexity (McMillan degree) of the solutions as parameter. This solution is nonrecursive. For

[1] We use the term *realization* to denote the problem which is usually referred to in the literature as *partial realization*.

[2] $\underline{n}$ denotes the set $\{1, 2, \cdots, n\}$; $e^{(\cdot)}$ denotes the exponential function.

details see Antoulas and Willems [7], Antoulas, Ball, Kang, and Willems [8]. Furthermore, in Antoulas [1], the complete recursive solution of the realization problem has been worked out in the generating system framework. It should be kept in mind that (as shown in Section II) the problems of *realization* and *rational interpolation* can be viewed as special cases of the general polynomial-exponential time series modeling introduced above.

2) Questions

The main question which arises is: what is the origin and system-theoretic interpretation of the generating system? Which property of the data guarantees its existence? In particular, while one is interested in single-input single-output system compatible with the data, where does the two-input two-output generating system come from? Alternatively, when can the solutions of a modeling problem be parametrized by means of linear fractions?

3) Answers

It is the thesis of this paper that these questions can be answered by adopting the *behavioral framework* to the modeling problem. It will be argued in Section III, that if we are interested in linear, time invariant models, the central object on the level of trajectories (behaviors) is the smallest linear and shift invariant cover $\mathbf{B}^*$ of $\mathbf{D}$, which we shall refer to as the *most powerful unfalsified model*, abbreviated *MPUM*, of the data $\mathbf{D}$. It follows that $\mathbf{B}^*$ is unique and all other linear and shift invariant covers $\mathbf{B}$ of $\mathbf{D}$ yield less powerful, unfalsified models of $\mathbf{D}$:

$$\mathbf{D} \subset \mathbf{B}^* \subset \mathbf{B}.$$

Part of the central construction, which will be discussed in detail below, is to express $\mathbf{B}^*$ as the *kernel* of an appropriate differential operator with constant coefficients $\Theta^*(d/dt)$:

$$\mathbf{B}^* = \ker \Theta^*, \quad \Theta^* \in \mathbf{k}^{g\times 2}[s], \quad g \leq 2.$$

Hence, $\Theta^*(d/dt)\mathbf{w} = 0$, for all $\mathbf{w} \in \mathbf{B}^*$, and consequently also for $\mathbf{w}_i \in \mathbf{D} \subset \mathbf{B}^*$, $i \in \underline{n}$. The following key result holds:

$$g = 2 \quad \text{and} \quad \det \Theta^* \neq 0 \Leftrightarrow \dim_{\mathbf{k}} \mathbf{B}^* < \infty.$$

Furthermore, any other (less powerful linear time invariant) model $\mathbf{B}$ can also be represented as the kernel of some differential operator $\Theta(d/dt)$:

$$\mathbf{B}^* = \ker \Theta^* \subset \ker \Theta =: \mathbf{B}.$$

The above inclusion implies the existence of a differential operator Γ such that

$$\Theta = \Gamma\Theta^* \qquad \text{where } \Theta, \Gamma \in \mathbf{k}^{g\times 2}[s],\ g = 1 \text{ or } 2.$$

The converse is also true, i.e., for every Γ as above we have

$$\mathbf{B}^* = \ker \Theta^* \subset \ker \Gamma\Theta^* =: \mathbf{B}.$$

This shows that Θ^* is indeed a generating system. In other words, *the existence of a finite dimensional linear time invariant cover for the data set* $\mathbf{D}$ *is equivalent to the existence of a two-input two-output generating system* Θ^* (also called a *two-port*; see Section X for details). Recall that the generating system is completely determined from the data, while the terminating system Γ is completely arbitrary.

Furthermore, the fact that in the above parametrization Γ is arbitrary yields, at least conceptually, the solution to the recursive modeling problem *for free*. In particular, if $\mathbf{D} = \mathbf{D}_1 \cup \mathbf{D}_2$ we define Θ_1^* as a generating system for the data set $\mathbf{D}_1$ and Θ_2^* as a generating system for the modified data set $\Theta_1^*(d/dt)\mathbf{D}_2$. The cascade $\Theta^* := \Theta_2^*\Theta_1^*$ of the two generating systems Θ_1^* and Θ_2^*, provides a generating system for $\mathbf{D}$.

4) Summary of Contents

In Section II the various problems which can be cast as special cases of the polynomial-exponential framework are discussed. Section III provides a comprehensive description of the behavioral framework. Only the ingredients needed for the solution of the present problem are discussed. They include the definition of linear systems, the concept of the most powerful unfalsified model (MPUM), behavioral equation representations of the MPUM, I/O systems, model complexity and model controllability. The section concludes with a precise statement of the polynomial-exponential modeling problem. Section IV is devoted to the construction of the most powerful unfalsified model $\mathbf{B}^*$, which as we already mentioned, is the smallest linear shift invariant cover of the data set $\mathbf{D}$. A basis for $\mathbf{B}^*$ is explicitely given in terms of the elements $\mathbf{D}$ (Theorem 4.4).

The next three sections are concerned with equation representations of the MPUM $\mathbf{B}^*$. In Section V, by inspection of $\mathbf{D}$, constant matrices H, F parametrizing $\mathbf{B}^*$ are written down. The *state variable* (SV) equation representation asserts that $\mathbf{w} \in \mathbf{B}^*$ if, and only if, there exists a state trajectory of appropriate dimension $\mathbf{x}$, such that $(d/dt)\mathbf{x} = F\mathbf{x}$, $\mathbf{w} = H\mathbf{x}$ (Theorem 5.5, Corollary 5.6). The elimination of the state $\mathbf{x}$ from these equations yields the so-called *autoregressive* (AR) equation representation: $\mathbf{w} \in \mathbf{B}^*$ if, and only if, $\Theta^*(d/dt)\mathbf{w} = 0$, i.e., $\mathbf{B}^* = \ker \Theta^*$ (Theorem 6.4); Θ^* is what we have called above a *generating* system.

The next two sections are devoted to *recursive modeling* of the data set $\mathbf{D}$. Section XIII treats the recursive update of the MPUM $\mathbf{B}^*$, as well as the update of the corresponding SV and AR representations.

An important sub-class of models are the *controllable* ones. The remaining part of the paper is devoted to the study of these models. In Section VII, AR representations of controllable systems are described in terms of the Smith canonical form (Formula 7.1). As explained in Section III-E-2) controllable models can be equivalently represented as the image of an appropriate differential operator: $\mathbf{B}_{contr} = im\Psi_{contr}$, i.e., for every $\mathbf{w} \in \mathbf{B}_{contr}$, there exists an auxiliary vector trajectory $\mathbf{a}$ of appropriate dimension, such that $\mathbf{w} = \Psi_{contr}(d/dt)\mathbf{a}$. These are the so-called *moving average* (MA) equation representations (Formula 7.2).

The difficulty in dealing with controllable models lies in the fact that in general, there is *no* controllable MPUM (see Example 3.22). Instead, there is a family of controllable models having minimal complexity, abbreviated C-MCUM (controllable minimal-complexity unfalsified models), defined in (3.23a). In order to characterize this family we need 2 ingredients: a) the *invariant factors* of any generating system Θ^*, and b) the *row degrees* of one row reduced Θ^*. The characterization is given recursively. Hence, the recursive update of the invariant factors and the row indices is discussed in Section VIII Theorem 8.13, Corollary 8.14.

Section IX discusses the recursive update of the family of minimal-complexity controllable models. This is accomplished in two steps. *First*, given *one* controllable model of minimal complexity, a parametization of *all* such models is achieved. This parametrization turns out to be affine (Theorem 9.9 of Section IX-A). *Second*, the update of *one* controllable model of minimal complexity is worked out in Section IX-B. The main result is Theorem 9.18, which gives the simultaneous update of the MPUM and of a C-MCUM.

This result is of technical nature. In order to make it as transparent as possible, we list the ingredients which are involved: a) a row reduced (with order row degrees) generating system at step n, denoted by Θ^*, and a controllable model of minimal complexity, consisting of certain rows of Θ^* [indexed by the set **I** defined in (9.2)]. b) the error ϵ_{n-1} defined by (8.8–9), and the corresponding error generating system Γ_{n+1} defined by (8.11). c) the indexes $r \in \underline{q}$ and $l, k \in \underline{q}$ defined by (8.10) and (9.11c, d), respectively, and the resulting matrix A defined by (9.17a–c). The update is then given explicitely by formulae (9.19).

Section X discusses the representation of the above results in terms of cascade interconnections of I/O systems and (where appropriate) of linear fractions (Figs. 1–3). We conclude with examples which illustrate the main features of the theory (Section XI).

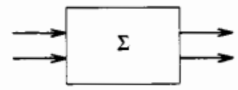

Fig. 1.

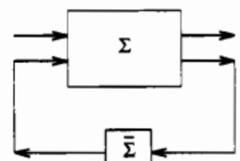

Fig. 2.

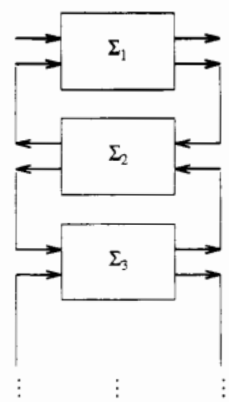

Fig. 3.

II. Related Problems

The solution of a number of important problems can be obtained as a corollary to the solution of the *main problem* defined in Section I-B and I-C. These problems are discussed below.

A) Finite Observation Interval. The solution of the *main problem* remains valid if the interval of observation, i.e., the domain of $\mathbf{w}_i$, is replaced by any nonzero finite interval $\mathbf{T} := [a, b] \subset \mathbf{R}$. □

B) Arbitrary Responses in Continuous-Time. It is well known that using Fourier analysis, a large class of vector-valued signals $\mathbf{w}: \mathbf{R} \to \mathbf{k}^q$, can be expressed as (possibly infinite) linear combinations of (complex) exponentials:

$$\mathbf{w} = \sum_i \alpha_i \mathbf{w}_i,$$

where $\mathbf{w}_i := p_i \exp_{\lambda_i}$, $\alpha_i \in \mathbf{k}$, $p_i \in \mathbf{k}^q$, $\lambda_i \neq \lambda_j$, $i \neq j$;

the summation might be discrete or continuous (integration). Since each $\mathbf{w}_i$ is linearly independent from the linear hull of all the remaining $\mathbf{w}_j$, $j \neq i$, a straightforward argument implies that a linear, time invariant system explains $\mathbf{w}$ if, and only if, it explains each $\mathbf{w}_i$ individually. We thus conclude that, at least *in principle*, modeling experiments involving fairly large classes of continuous-time signals can be reduced to *modeling experiments involving single-frequency exponential signals*. □

C) The Rational Interpolation Problem. Consider a linear system with unknown rational transfer function Z of size $p \times m$. Let the data supplied be samples of the *frequency response*, i.e., the values of this transfer function, and the values of a number of consecutive derivatives thereof, at given frequencies:

$$\lambda_i;\ Z(\lambda_i), Z^{(1)}(\lambda_i), \cdots, Z^{(\kappa_i - 1)}(\lambda_i); \qquad i \in \underline{n}; \quad (2.1)$$

$Z^{(r)}(\lambda)$ denotes the rth derivative of Z evaluated at λ. The problem of recovering all Z compatible with the above data, is often referred to as *rational interpolation*. For simplicity, let $m = p = 1$, i.e., $q = 2$. With $\mathbf{w}_i := \begin{pmatrix} \mathbf{u}_i \\ \mathbf{y}_i \end{pmatrix}$, we define the polynomial-exponential time series $\mathbf{w}_i: \mathbf{R} \to \mathbf{k}^2$, where

$$\mathbf{w}_i(t) := \exp_{\lambda_i}(t)\left\{\begin{pmatrix} 1 \\ Z(\lambda_i) \end{pmatrix}\frac{t^{\kappa_i-1}}{(\kappa_i-1)!} + \cdots + \begin{pmatrix} 0 \\ \frac{Z^{(j)}(\lambda_i)}{j!} \end{pmatrix}\frac{t^{\kappa_i-j-1}}{(\kappa_i-j-1)!} + \cdots + \begin{pmatrix} 0 \\ \frac{Z^{(\kappa_i-1)}(\lambda_i)}{(\kappa_i-1)!} \end{pmatrix}\right\}.$$

A straightforward argument involving the Taylor series expansion of $(s-\lambda_i)^{\kappa}\mathbf{W}_i(s)$ around $s = \lambda_i$, where $\mathbf{W}_i(s)$ denotes the Laplace transform of $\mathbf{w}_i$, shows that the transfer function of any linear system compatible with the input–output experiment which is defined by $\mathbf{w}_i$, will satisfy (2.1). Notice moreover, that because of the 0's appearing in the upper row of $\mathbf{w}_i$, rational interpolation is a *special case* of the general problem of polynomial-exponential time series modeling. □

D) The Realization or Impulse Response Modeling Problem. As in C), for simplicity, consider single–input, single–output systems. Given the scalars $a_0, a_1, \cdots, a_{N-1}$, determine all rational functions Z whose behavior at infinity (i.e., formal power series) is:

$$Z(s) = a_0 + a_1 s^{-1} + \cdots + a_{N-1}s^{-N+1} + \cdots.$$

Sometimes, the a_i are referred to as *Markov parameters*. By introducing s^{-1} as the new variable, the behavior at infinity is transformed into the behavior at zero:

$$\bar{Z}(s) := Z(s^{-1}) = a_0 + a_1 s + \cdots + a_{N-1}s^{N-1} + \cdots.$$

Consequently, $a_i = \bar{Z}^{(i)}(0)/i!$, i.e., the realization problem is equivalent to a rational interpolation problem where all the data are provided at zero. From C), the corresponding time series $\mathbf{w}: \mathbf{R} \to \mathbf{K}^2$, defined by

$$\mathbf{w}(t) := \begin{pmatrix} 1 \\ a_0 \end{pmatrix}\frac{t^{N-1}}{(N-1)!} + \cdots + \begin{pmatrix} 0 \\ a_j \end{pmatrix}\frac{t^{N-j-1}}{(N-j-1)!} + \cdots + \begin{pmatrix} 0 \\ a_{N-1} \end{pmatrix}, \quad (2.2)$$

is thus purely polynomial. □

E) Transient and Steady-State Response. Experiments which involve both transient and steady-state information, can be cast into the above framework. Here is a simple example. Suppose that the input of a (scalar) linear system is $\mathbf{u} = \exp_\lambda$, $t \geq 0$, while the resulting output is $\mathbf{y} = c_\lambda \exp_\lambda + c_\mu \exp_\mu$, $t \geq 0$. The time series corresponding to this experiment is

$$\mathbf{w} = \begin{pmatrix} 1 \\ c_\lambda \end{pmatrix}\exp_\lambda + \begin{pmatrix} 0 \\ c_\mu \end{pmatrix}\exp_\mu, \qquad t \geq 0.$$

Assume that $\lambda \neq \mu$; due to the linear independence of the two summands of $\mathbf{w}$, each one can be modeled *individually*. The first term of $\mathbf{w}$ forces the value of the transfer function of the underlying system at the frequency λ to be equal to c_λ, while the second term says that μ is an eigenfrequency of the system, i.e., a pole of the transfer function. Since for modeling purpose, this term is equivalent to $\begin{pmatrix} 0 \\ 1 \end{pmatrix}\exp_\mu$, the coefficient c_μ does *not* enter into the determination of the model. Indeed, this constant is due to the *initial condition* of the system which we are modeling. □

Remarks 2.3:

a) The problems described in C), D), and E) can be formulated for multiinput, multioutput systems at the expense of more involved notation. The essence however remains: many exact modeling problems, including the widely known problems of rational interpolation and realization, can be cast into the framework of polynomial-exponential time series modeling.

b) The discrete-time version of the realization problem described in D) has the following interpretation in terms of input–output experiments. Let the input $\mathbf{u}$ be an impulse, i.e., $\mathbf{u}(0) = 1$ and $\mathbf{u}(t) = 0$, for $t \neq 0$. The corresponding output being $\mathbf{y}(t) = a_t$, for $t \geq 0$, and $\mathbf{y}(t) = 0$, $t < 0$, we are seeking all linear, time invariant models having the above impulse response up to time $t = N - 1$. Thus, in D), this discrete-time problem, defined on a time interval which depends on the number of data points, is transformed into a continuous-time problem defined on a *fixed-length* time interval. This transformation is applied to a more general discrete-time problem in G) below. □

F) Discrete-Time Observation Interval. The main problem can be formulated for discrete-time polynomial-exponential observations, i.e., in (1.1) $\mathbf{T} = \mathbf{Z}$. The solution of the continuous-time problem $\mathbf{T} = \mathbf{R}$, applies mutatis mutandis to this case. □

G) Arbitrary Responses in Discrete-Time. A problem of great significance for applications is that of modeling on the basis of arbitrary discrete-time experiments: given an observed time series $\mathbf{w}: \mathbf{Z} \to \mathbf{k}^q$, we are looking for all discrete-time, linear, time invariant models explaining[3] it.

This problem can be solved from first principles as shown in Antoulas [11]. It can also be solved by transforming it to a continuous-time problem as follows. Assume, without loss of generality (cf. E) above), that the experiment has started at time $t = 0$, i.e., $\mathbf{w}(t) = 0$, $t < 0$. Define the following continuous-time time series $\mathbf{v}_N: \mathbf{R} \to \mathbf{k}^q$ where

$$\mathbf{v}_N(t) := \mathbf{w}(0)\frac{t^{N-1}}{(N-1)!} + \cdots + \mathbf{w}(j)\frac{t^{N-j-1}}{(N-j-1)!} + \cdots + \mathbf{w}(N-1).$$

Denote the family of discrete-time, linear, time invariant systems Σ_N^{disc}, which explain[3] $\mathbf{w}$ up to time $N-1$, and the

[3] A model "explains" a time series, if the time series could have been generated by the model in question.

family of continuous-time, linear, time invariant systems Σ_N^{cont}, which explain[3] $\mathbf{v}_N$, respectively by: $\mathbf{F}_N^{disc} := \{\Sigma_n^{disc}\}$, $\mathbf{F}_N^{cont} := \{\Sigma_N^{cont}\}$. Notice that the elements of $\mathbf{F}_N^{disc}$ are defined in terms of the forward shift operator σ^{-1}, while those of $\mathbf{F}_N^{cont}$ in terms of the derivative operator (d/dt). There is a one-to-one correspondence $\phi: \mathbf{F}_N^{cont} \to \mathbf{F}_N^{disc}$, between these two families. It is obtained by replacing the *derivative* operator by the *forward shift* operator: $(d/dt) \to \sigma^{-1}$. This follows as a straightforward generalization of the discrete-time realization problem discussed above (see (2.2) and Remark 2.3-b). For an illustration, see Example A of Section XI.

In connection with the single, discrete time series $\mathbf{w}$, a recursive modeling problem can be defined with respect to *time*. Given a discrete-time model explaining $\mathbf{w}$ up to time t, call it Σ_t, and $\mathbf{w}(t+1)$, compute Σ_{t+1}. Clearly, using the above equivalence, this problem, defined on a variable-length time interval, can be transformed into a problem of recursive modeling of the set of continuous-time time series: $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_N$, defined over the fixed-length interval $\mathbf{R}$, recursivity being with respect to the index i, and not with respect to time t. □

Remark 2.4: Realization was the first exact modeling problem to be studied in the literature. It was formally introduced in the 60's (see Kalman, Falb, and Arbib [21]) and eventually two approaches crystallized: the *state space* and the *polynomial* (see Fuhrmann [16] for an overview of the interplay between these two approaches in linear system theory). The state space method uses the Hankel matrix as main tool; see, e.g., Bosgra [13], Antoulas, Matsuo, and Yamamoto [10] for a recent overview, and Gohberg, Kaashoek, and Lerer [17] for a generalization. The polynomial approach has the Euclidean division algorithm as focal point; see, e.g., Kalman [20], Fuhrmann [15], Antoulas [1], van Barel and Bultheel [24]. Actually, Antoulas [1] presents the presents the complete theory of recursive realization for multiinput, multioutput systems, and as already mentioned in the overview, anticipates the recursiveness results presented in Sections VIII and IX.

The problem of *rational interpolation* has a long history. It was only recently recognized however, as an exact modeling problem which generalizes the realization problem. Again, one can distinguish two approaches: state space and polynomial. The generalization of the state space framework from the realization to the rational interpolation problem is due to Antoulas and Anderson [2], [3], [6] and Anderson and Antoulas [5]. Therein, the Löwner matrix replaces and generalizes the Hankel matrix as the main tool. A polynomial approach to rational interpolation, which can be viewed as a special case of the present theory, was put forward in Antoulas and Willems [7] and Antoulas, Ball, Kang, and Willems [8]. For a recent, general account on the rational interpolation problem, see Ball, Gohberg, and Rodman [12]. □

III. The Behavioral Framework

This section is devoted to an overview of the behavioral framework. For further details and proofs, the original sources, viz. Willems [25], [26], [28], [29], are to be consulted. The behavioral framework is built around the triptych:

behavior—behavioral equations—latent variables.

The brief discussion of these concepts that follows, is tailored to the needs of the modeling problem under investigation. It should be noted that a number of the aspects discussed below can be found in earlier work, notably in electrical network theory, and in the work of Belevich and Rosenbrock. We refer to the original sources just mentioned for details.

A. Linear Systems

A *dynamical system* Σ is an object composed of three quantities:

$$\Sigma := (\mathbf{T}, \mathbf{W}, \mathbf{B}); \tag{3.1}$$

$\mathbf{T} \subseteq \mathbf{R}$ is the *time axis*; $\mathbf{W}$ is the *signal space*, i.e., the space in which the external or manifest system trajectories $\mathbf{w}$ take their values; and $\mathbf{B}$ is a subset of $\mathbf{W}^{\mathbf{T}}$, the space of all time-trajectories, i.e., maps $\mathbf{w}: \mathbf{T} \to \mathbf{W}$ from the time axis to the signal space. The collection of trajectories $\mathbf{B}$ is called the **behavior** of Σ. In the sequel we will consider continuous-time systems with

$$\mathbf{T} = \mathbf{R} \quad \text{and} \quad \mathbf{W} = \mathbf{k}^q, q \geq 1.^4$$

For systems with m inputs and p outputs, $q = m + p$. Hence, $\mathbf{B} \subset (\mathbf{k}^q)^{\mathbf{R}}$ is a collection of q-vector valued time series.[5] The properties of linearity and time invariance of Σ, are reflected in a natural way on the behavior $\mathbf{B}$. The system $\Sigma = (\mathbf{R}, \mathbf{k}^q, \mathbf{B})$ is *linear* iff $\mathbf{B}$ is a linear subspace of $(\mathbf{k}^q)^{\mathbf{R}}$, i.e.,

$$(\mathbf{w}_1, \mathbf{w}_2 \in \mathbf{B} \text{ and } \alpha_1, \alpha_2 \in \mathbf{k}) \Rightarrow (\alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2 \in \mathbf{B});$$

time invariant if

$$(\mathbf{w} \in \mathbf{B}) \Leftrightarrow (\sigma\mathbf{w} \in \mathbf{B}),$$

where σ denotes the backward shift: $(\sigma\mathbf{w})(t) := \mathbf{w}(t+1)$.

The property of finite dimensionality is reflected in the behavior in a more involved way. Σ is *finite-dimensional*, if $\mathbf{B}$ is *differential* or *instantaneously specified*, i.e., if $\mathbf{B}$ can be described as the solution set of a system of differential equations. For the definition and for a discussion of the concept of differential or instantaneously specified behaviors, see Willems [29, p. 279]. For our purposes, we only need the fact that this property is equivalent with the representability of $\mathbf{B}$ as the kernel of a polynomial operator in the differentiation operator (d/dt), in an appropriate sense, as described in the reference just mentioned.

In the sequel we will always be dealing with the family of *continuous-time, linear, time invariant, finite dimensional systems*, which will be denoted by

$$\mathbf{L}^q := \{\Sigma = (\mathbf{R}, \mathbf{k}^q, \mathbf{B}): \mathbf{B} \text{ is linear, shift invariant, differential}\}; \tag{3.2}$$

[4] Recall that $\mathbf{k}$ denotes $\mathbf{R}$ or $\mathbf{C}$.

[5] The term *time series* is used here as an alternative to *time function*.

we will refer to the elements of $\mathbf{L}^q$ simply as **linear** systems.

B. Modeling and the MPUM

On the level of behaviors, the *modeling problem* can be formalized as follows Let a finite set of trajectories, called the *data* set and denoted by $\mathbf{D}$, be observed and measured:

$$\mathbf{D} := \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n\} \qquad \text{with } \mathbf{w}_i: \mathbf{R} \to \mathbf{k}^q, \quad i = 1, 2, \cdots, n.$$

Thus, $\mathbf{D} \subset (\mathbf{k}^q)^{\mathbf{R}}$. The system $\Sigma = (\mathbf{R}, \mathbf{k}^q, \mathbf{B}) \in \mathbf{L}^q$ will be called an *unfalsified model* of $\mathbf{D}$ iff

$$\mathbf{D} \subseteq \mathbf{B}.$$

Let Σ_1, Σ_2 be unfalsified models of $\mathbf{D}$. We will call Σ_1 a subsystem of Σ_2, denoted $\Sigma_1 \subseteq \Sigma_2$, if $\mathbf{B}_1 \subseteq \mathbf{B}_2$. Equivalently, we will call Σ_1 *more powerful* than Σ_2. This has the following intuitive interpretation in the context of modeling: the *more* powerful a mathematical model is, i.e., the *less* its behavior explains, the *more predictive power* it has.

This brings us to the concept of the *most powerful unfalsified model*, abbreviated MPUM. We will call $\Sigma^* := (\mathbf{R}, \mathbf{k}^q, \mathbf{B}^*)$[6] the *most powerful model* in $\mathbf{L}^q$, *unfalsified* by $\mathbf{D}$, if

$$(\Sigma \in \mathbf{L}^q \text{ and } \Sigma \text{ unfalsified by } \mathbf{D}) \Rightarrow (\Sigma^* \subset \Sigma). \quad (3.3)$$

In Willems [26] it is shown that given $\mathbf{D}$, the MPUM in the class $\mathbf{L}^q$ *exists*, and is, therefore, unique. Thus, the MPUM explains the given data set and as little else as necessary so as to obtain a model in the desired class, the class of linear systems in this case.

Depending on the choice of the model class the MPUM may *not* exist. In such a case the weaker concept of *undominated* model is used instead. We will call Σ an *undominated* model in some given class $\mathbf{M}$, *unfalsified* by $\mathbf{D}$, if $\Sigma \in \mathbf{M}$, and if

$$(\Sigma' \in \mathbf{M}, \Sigma' \text{ unfalsified by } \mathbf{D}, \text{ and } \Sigma' \subset \Sigma) \Rightarrow (\Sigma' = \Sigma). \quad (3.4)$$

It will be shown later that the class $\mathbf{M} := \mathbf{L}^q_{contr} \subset \mathbf{L}^q$, of linear controllable models of $\mathbf{D}$, is an example of a model class in which an MPUM does not exist in general. Clearly, the most powerful model, if it exists, is also undominated, but not vice versa.

For the problem at hand, the data set $\mathbf{D}$ is composed of the collection of *polynomial-exponential* time series defined in (1.1)–(1.3). In Section IV, the MPUM Σ^* of $\mathbf{D}$ in the class $\mathbf{L}^q$ is explicitely computed; its behavior $\mathbf{B}^*$ turns out to be *finite-dimensional*:

$$\dim_{\mathbf{k}} \mathbf{B}^* = N < \infty.$$

This implies that Σ^* is an *autonomous* system, i.e., a system whose behavior has the property that each trajectory in $\mathbf{B}^*$ is uniquely determined by its past:

$$\{T \in \mathbf{T}, \mathbf{w}_1, \mathbf{w}_2 \in \mathbf{B}^*; \mathbf{w}_1(t) = \mathbf{w}_2(t), t < T\} \Rightarrow \{\mathbf{w}_1 = \mathbf{w}_2\}. \quad (3.5)$$

For details on autonomous systems, see last paragraph of Section III-D-3).

C. Behavioral Equations

Most often, it is very convenient to represent the trajectories making up the behavior of a system, as solutions of appropriate equations. The second concept of our triptych is that of **behavioral equations**. The third are **latent variables.** Behavioral equations with latent variables involve the manifest variables $\mathbf{w}$. Three instances of behavioral equations, which will be used in the sequel, are: *state variable* (abbreviated SV) equations; *autoregressive* (abbreviated AR) equations; and *moving average* (abbreviated MA) equations. AR equations involve the manifest variables $\mathbf{w}$ only; SV equations, in addition to $\mathbf{w}$, make use of the latent variables $\mathbf{x}$, which can be assigned the property of state, and are called *state* variables; and MA equations make use of $\mathbf{w}$ and of the latent variables $\mathbf{a}$, called *auxiliary* or *driving* variables.

1) State Variable (SV) Behavioral Equations

For autonomous systems in $\mathbf{L}^2$, the general theory developed in Willems [25], implies the existence of a representation of Σ^* in terms of SV behavioral equations:

$$\frac{d}{dt}\mathbf{x} = F\mathbf{x}, \quad \mathbf{w} = H\mathbf{x}, \quad \mathbf{x} \in (\mathbf{k}^N)^{\mathbf{R}}, \quad \mathbf{w} \in (\mathbf{k}^q)^{\mathbf{R}}, \quad (3.6a)$$

for an appropriate (observable) pair $(H, F) \in \mathbf{k}^{q \times N} \times \mathbf{k}^{N \times N}$. The behavior can thus be represented in terms of all possible trajectories generated by the above equations:

$$\mathbf{B}^* = \mathbf{B}(H, F) := \{\mathbf{w}: (3.6a) \text{ is satisfied for some } \mathbf{x} \in (\mathbf{k}^N)^{\mathbf{R}}\}. \quad (3.6b)$$

Section V addresses the problem of constructing SV behavioral equations for the data set $\mathbf{D}$, given by (1.1)–(1.3).

2) Autoregressive (AR) Behavioral Equations

By eliminating the state variable $\mathbf{x}$ from (3.6a) we obtain AR (autoregressive)[7] behavioral equations representing Σ^*. Let $\mathbf{k}[s]$ denote the ring of polynomials in the indeterminate s with coefficients in $\mathbf{k}$, and $\mathbf{k}^{n_1 \times n_2}[s]$ denote the $n_1 \times n_2$ polynomial matrices. The resulting AR equation has the form:

$$\Theta^*\left(\frac{d}{dt}\right)\mathbf{w} = 0, \quad \Theta^* \in \mathbf{k}^{q \times q}[s], \quad \det \Theta^* \neq 0. \quad (3.7a)$$

It relates the time series $\mathbf{w}: \mathbf{R} \to \mathbf{k}^q$ belonging to $\mathbf{B}$, to its derivatives. This equations can be written explicitely in terms of the coefficient matrices of Θ^*. Let

$$\Theta^*(s) := \Theta_L s^L + \Theta_{L-1} s^{L-1} + \cdots + \Theta_1 s + \Theta_0, \quad \Theta_i \in \mathbf{k}^{q \times q}. \quad (3.7b)$$

[6] Superscript * denotes either an MPUM-related quantity, or more generally, a minimal-complexity-related quantity.

[7] An *autoregressive* equation is literally an equation which regresses on itself. No stochastic connotation is implied.

Equation (3.7a) becomes:

$$\Theta_L \frac{d^L}{dt^L}\mathbf{w}(t) + \Theta_{L-1}\frac{d^{L-1}}{dt^{L-1}}\mathbf{w}(t) + \cdots + \Theta_1 \frac{d}{dt}\mathbf{w}(t) + \Theta_0 \mathbf{w}(t) = 0, \qquad t \in \mathbf{R}. \quad (3.7c)$$

The details of how Θ^* is obtained from (3.6a) can be found in Section VI. Every differential operator which annihilates the behavior is called **annihilating.** Equation (3.7a) is thus sometimes referred to as an *annihilating* behavioral equation. Let $C^\infty(X,Y)$ denote the set of all infinitely differentiable maps from X to Y. $\Theta^*(d/dt)$ can be interpreted as a map from $C^\infty(\mathbf{R},\mathbf{k}^q)$ to $C^\infty(\mathbf{R},\mathbf{k}^q)$. The following representation of the behavior $\mathbf{B}^*$ is thus obtained:

$$\mathbf{B}^* = \ker \Theta^*\left(\frac{d}{dt}\right) := \{\mathbf{w}: \mathbf{R} \to \mathbf{k}^q, \text{ with (3.7a) satisfied}\}. \quad (3.8)$$

In the sequel we will use the following notation for systems defined as above:

$$\Sigma(\Theta) := (\mathbf{R}, \mathbf{k}^q, \mathbf{B}(\Theta)) \qquad \text{where } \mathbf{B}(\Theta) := \ker \Theta\left(\frac{d}{dt}\right). \quad (3.9)$$

Thus, from (3.8), $\Sigma^* = \Sigma(\Theta^*)$. The fact that we are dealing with an autonomous system is reflected in the nonsingularity of Θ^*, i.e., $\det \Theta^* \neq 0$ [see last paragraph of Section III-D-3)]. The spaces $\mathbf{B}(\Theta)$ defined above, satisfy the following properties.

Proposition 3.10: Consider the maps $\Theta_i(d/dt)$: $C^\infty(\mathbf{R},\mathbf{k}^q) \to C^\infty(\mathbf{R},\mathbf{k}^{n_i})$, where $\Theta_i \in \mathbf{k}^{n_i \times q}[s]$, $i = 1, 2$.

a) Let $\Theta_2 = \Gamma\Theta_1$, $\Gamma \in \mathbf{k}^{n_2 \times n_1}[s]$.
 a1) $\mathbf{B}(\Theta_1) \subseteq \mathbf{B}(\Theta_2)$.
 a2) Γ is unimodular if, and only if, $\mathbf{B}(\Gamma) = 0$.
 a3) If Γ is unimodular, $\mathbf{B}(\Theta_1) = \mathbf{B}(\Theta_2)$. Conversely, if $\mathbf{B}(\Theta_1) = \mathbf{B}(\Theta_2)$ and Θ_1 has full row rank, Γ is unimodular.

b) $\mathbf{B}(\Theta_1) \cap \mathbf{B}(\Theta_2) = \mathbf{B}(\tilde{\Theta})$, where $\tilde{\Theta} \in \mathbf{k}^{\hat{q}\times q}[s]$, $\hat{q} \le q$, is the greatest common right divisor of the given polynomial matrices: $\tilde{\Theta} = gcrd(\Theta_1, \Theta_2)$.

c) Conversely to a1), $\mathbf{B}(\Theta_1) \subseteq \mathbf{B}(\Theta_2)$ implies the existence of $\Gamma \in \mathbf{k}^{n_2 \times n_1}[s]$, such that $\Theta_2 = \Gamma\Theta_1$.

d) If $\det \Theta_2 \neq 0$, $\mathbf{B}(\Theta_1\Theta_2) = \Theta_2^{-1}\mathbf{B}(\Theta_1)$.

e) If $n_1 = n_2 = q$, $\mathbf{B}(\Theta_1) + \mathbf{B}(\Theta_2) = \mathbf{B}(\hat{\Theta})$, where $\hat{\Theta} \in \mathbf{k}^{q \times q}[s]$ is the least common left multiple of the given polynomial matrices: $\hat{\Theta} = lclm(\Theta_1, \Theta_2)$.

Proof: a1) follows trivially. If Γ is unimodular, by a1), $\Gamma\Gamma^{-1} = I$ implies $\mathbf{B}(\Gamma) \subset \mathbf{B}(I) = 0$. Conversely, if Γ is not unimodular, clearly $\mathbf{B}(\Gamma) \neq 0$. This proves a2). Since $\Theta_1 = \Gamma^{-1}\Theta_2$, with Γ unimodular, a1) implies $\mathbf{B}(\Theta_2) \subseteq \mathbf{B}(\Theta_1)$; consequently

$$\mathbf{B}(\Gamma\Theta_1) = \mathbf{B}(\Theta_1) = \mathbf{B}(\Theta_2).$$

If Γ is not unimodular, by a2), there exists $z \neq 0$ such that $\Gamma z = 0$; since Θ_1 has full row rank, there exists y such that $z = \Theta_1 y$; the implication is that in this case the inclusion $\mathbf{B}(\Theta_2) \subset \mathbf{B}(\Theta_1)$ is strict, proving a3).

Since Θ_1, Θ_2 and their greatest common right divisor $\tilde{\Theta}$, satisfy $\Gamma\begin{pmatrix}\Theta_1\\ \Theta_2\end{pmatrix} = \begin{pmatrix}\tilde{\Theta}\\ 0\end{pmatrix}$, for some unimodular Γ, the validity of b) follows:

$$\mathbf{B}(\Theta_1) \cap \mathbf{B}(\Theta_2) = \mathbf{B}\begin{pmatrix}\Theta_1\\ \Theta_2\end{pmatrix} = \mathbf{B}\begin{pmatrix}\tilde{\Theta}\\ 0\end{pmatrix} = \mathbf{B}(\tilde{\Theta}).$$

Consequently, $\mathbf{B}(\Theta_1) \subseteq \mathbf{B}(\Theta_2)$ implies $\mathbf{B}(\tilde{\Theta}) = \mathbf{B}(\Theta_1)$, where $\Theta_1 = \Phi_1\tilde{\Theta}$; the above equality of behaviors implies by a3) the unimodularity of Φ_1, since $\tilde{\Theta}$ can always be chosen to have full row rank. We conclude that Θ_1 divides $\tilde{\Theta}$ which in turn divides Θ_2, as required; part c) is thus proved.

Part d) follows readily, provided that the inverse image of Θ_2 is well defined. This is indeed the case because both the domain and the range of the derivative operators $\Theta_i(d/dt)$, consist of C^∞ functions.

Since by construction Θ_i is a right divisor of $\hat{\Theta}$, $i = 1, 2$, a1) implies $\mathbf{B}(\Theta_i) \subset \mathbf{B}(\hat{\Theta})$, and hence

$$\mathbf{B}(\Theta_1) + \mathbf{B}(\Theta_2) \subseteq \mathbf{B}(\hat{\Theta}).$$

To prove the converse inclusion, let $\Theta_i = \Phi_i\tilde{\Theta}$, $i = 1, 2$, where $\tilde{\Theta} := gcrd(\Theta_1, \Theta_2)$. Let furthermore, the following generalized Bezout identity hold (see, e.g., Kailath [19, ch. 6.3]):

$$\begin{pmatrix}\Psi_1 & \Psi_2\\ A & -B\end{pmatrix}\begin{pmatrix}C & \Phi_1\\ D & -\Phi_2\end{pmatrix} = \begin{pmatrix}I_q & 0\\ 0 & I_q\end{pmatrix}.$$

It follows that the least common left multiple of Φ_1, Φ_2, is $\Phi_{12} := \Psi_1\Phi_1 = \Psi_2\Phi_2$, and that of Θ_1, Θ_2 is $\hat{\Theta} := \Phi_{12}\tilde{\Theta}$. The Bezout identity implies:

$$I_q = A\Phi_1 + B\Phi_2, \quad \Phi_2 A = D\Psi_1, \quad \Phi_1 B = C\Psi_2.$$

Thus, every $\mathbf{w}$ can be decomposed in the sum $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1 := A\Phi_1\mathbf{w}$ and $\mathbf{w}_2 := B\Phi_2\mathbf{w}$. If $\mathbf{w} \in \mathbf{B}(\Phi_{12})$, we have $\Phi_{12}\mathbf{w} = \Psi_1\Phi_1\mathbf{w} = \Psi_2\Phi_2\mathbf{w} = 0$; hence $\Phi_2\mathbf{w}_1 = \Phi_2 A\Phi_1\mathbf{w} = D\Psi_1\Phi_1\mathbf{w} = 0$ and $\Phi_1\mathbf{w}_2 = \Phi_1 B\Phi_2\mathbf{w} = C\Psi_2\Phi_2\mathbf{w} = 0$. Consequently $\mathbf{w} \in \mathbf{B}(\Phi_{12})$ implies $\mathbf{w}_2 \in \mathbf{B}(\Phi_1)$ and $\mathbf{w}_1 \in \mathbf{B}(\Phi_2)$, i.e.,

$$\mathbf{B}(\Phi_{12}) \subseteq \mathbf{B}(\Phi_1) + \mathbf{B}(\Phi_2).$$

Finally, because of d), $\tilde{\Theta}^{-1}\mathbf{B}(\Phi_i) = \mathbf{B}(\Theta_i)$ and $\tilde{\Theta}^{-1}\mathbf{B}(\Phi_{12}) = \mathbf{B}(\hat{\Theta})$; these relationships imply the desired inverse inclusion

$$\mathbf{B}(\hat{\Theta}) \subseteq \mathbf{B}(\Theta_1) + \mathbf{B}(\Theta_2),$$

which completes the proof of e). □

Given Θ^*, it is now straightforward to obtain AR equation representations for *all* other unfalsified models of $\mathbf{D}$ in $\mathbf{L}^q$. It follows from Definition 3.3 of the MPUM that, $\Sigma = (\mathbf{R}, \mathbf{k}^q, \mathbf{B}) \in \mathbf{L}^q$, is a linear unfalsified model of $\mathbf{D}$ if, and only if, $\mathbf{B}^* \subset \mathbf{B}$. From a) above, using notation

(3.9), follows that $\Sigma(\Theta)$, $\Theta \in \mathbf{k}^{g\times q}[s]$, is an unfalsified linear model of **D** if, and only if,

$$\Theta = \Gamma\Theta^* \quad \text{for some } \Gamma \in \mathbf{k}^{g\times q}[s], \text{ where } w.l.o.g.\ g \le q. \tag{3.11}$$

The corresponding behavioral equation representation of $\Sigma(\Theta) \supset \Sigma(\Theta^*)$ is:

$$\Theta\left(\frac{d}{dt}\right)\mathbf{w} = 0, \quad \Theta \in \mathbf{k}^{g\times q}[s] \qquad \text{for } \mathbf{w} \in \mathbf{B}(\Theta) \supset \mathbf{B}(\Theta^*). \tag{3.12}$$

Since by (3.11) every model of the data can be obtained from Θ^*, this latter quantity is sometimes called a *generating system* for the models of **D**.

D. The Complexity

From (3.11) it follows that $\mathbf{L}^q$ is parametrized by the set of matrix polynomials having q columns. The number of rows in (3.12) representing $\Sigma \in \mathbf{L}^q$ will depend on the particular system, and on the particular matrix Θ chosen to represent Σ. The AR behavioral equation representation (3.12) will be called *minimal* if

$$(\Theta' \in \mathbf{k}^{g'\times q}[s] \text{ and } \Sigma(\Theta) = \Sigma(\Theta')) \Rightarrow (g' \ge g),$$

i.e., Θ has the least number of rows. It is easy to prove that (3.12) is minimal if, and only if, the polynomial matrix Θ has full row rank (meaning that it contains a $g \times g$ submatrix, whose determinant is not the zero polynomial).

Obviously, any system $\Sigma \in \mathbf{L}^q$ admits a minimal representation. It follows that the number of rows in a minimal representation depends only on $\Sigma \in \mathbf{L}^q$, and not on the particular representation as a set of differential equations. This allows the association of an integer invariant p to each element of $\mathbf{L}^q$. In other words we have a map $p: \mathbf{L}^q \to \{0, 1, \cdots, q\}$, such that $p(\Sigma)$ equals the number of rows of a minimal behavioral equation representation (3.12) for Σ. In fact, if $\Sigma = \Sigma(\Theta)$, then $p(\Sigma) = rank\Theta$, with the rank as defined above.

It is possible to classify all minimal AR equation representations in terms of one. Indeed, $\Sigma(\Theta_1)$ and $\Sigma(\Theta_2)$ are both minimal representations of the same dynamical system if, and only if, there exists a polynomial unimodular matrix Γ of size $p(\Sigma)$, such that $\Theta_1 = \Gamma\Theta_2$.

1) I / O Systems

Assume that in (3.12) the signal **w** is partitioned as

$$\mathbf{w} = \begin{pmatrix}\mathbf{u}\\ \mathbf{y}\end{pmatrix} \qquad \text{with } \mathbf{u}: \mathbf{R} \to \mathbf{k}^m, \mathbf{y}: \mathbf{R} \to \mathbf{k}^p \text{ and } m + p = q.$$

Then, if we partition Θ accordingly, i.e., $\Theta := (\mathbf{Q}\ -\mathbf{T})$, we can write

$$\mathbf{T}\left(\frac{d}{dt}\right)\mathbf{y} = \mathbf{Q}\left(\frac{d}{dt}\right)\mathbf{u}$$
$$\text{with } \mathbf{T} \in \mathbf{k}^{g\times p}[s] \text{ and } \mathbf{Q} \in \mathbf{k}^{q\times m}[s]^8. \tag{3.13a}$$

[8] Recall that s is an indeterminate and *not* the Laplace variable.

Assume in addition that

$$g = p \quad \text{and} \quad \det \mathbf{T} \ne 0. \tag{3.13b}$$

Then it can be shown (see Willems [29]) that for all $\mathbf{u}: \mathbf{R} \to \mathbf{k}^m$, sufficiently smooth, there exists a $\mathbf{y}: \mathbf{R} \to \mathbf{k}^p$ such that (3.13a, b) holds. Consequently, we say that **u** is *free* or an *input*. Moreover, for any given sufficiently smooth $\mathbf{u}: \mathbf{R} \to \mathbf{k}^m$ the set of time series $\mathbf{y}: \mathbf{R} \to \mathbf{k}^p$, such that (3.13a) is satisfied will be a finite dimensional linear variety. Consequently, we say that **y** is *bound* or an *output*. We will say that (3.13a, b) defines an *input–output system*, abbreviated *I/O system* and denoted by $\Sigma_{I/O} := (\mathbf{R}, \mathbf{k}^m \times \mathbf{k}^p, \mathbf{B}_{I/O})$ where

$$\mathbf{B}_{I/O} := \mathbf{B}(\mathbf{Q}\quad -\mathbf{T})$$
$$:= \{(\mathbf{u},\mathbf{y}): \text{such that (3.13a, b) is satisfied}\}. \tag{3.14}$$

If in addition, the matrix of rational functions, called the *transfer function*, defined by:

$$Z := \mathbf{T}^{-1}\mathbf{Q} \in \mathbf{k}^{p\times m}(s)^8, \tag{3.15a}$$

is *proper*, the output **y** is at least as smooth as the input **u**, and we will refer to a *smooth I/O system*. This properness condition can be expressed in different terms. Due to the fact that $\mathbf{B}(\Theta) = \mathbf{B}(\Gamma\Theta)$, for any unimodular Γ, we may assume without loss of generality, that the polynomial matrix $\Theta = (\mathbf{Q}\quad -\mathbf{T})$ is *row reduced*. Let $[M]_{hr}$ denote the constant matrix which is composed of the coefficients of the highest power of each row of the polynomial matrix M. It can be shown (see e.g., Kailath [19]) that the properness of (3.15a), assuming **T** to be row reduced, is equivalent to the condition

$$\text{rank}\,[\mathbf{Q}\quad -\mathbf{T}]_{hr} = \text{rank}\,[\mathbf{T}]_{hr} = p. \tag{3.15b}$$

In Willems [29] it is shown that every $\Sigma \in \mathbf{L}^q$ induces a smooth I/O-system. More precisely, for any $\Sigma \in \mathbf{L}^q$ there will exist a $q \times q$ permutation matrix Π such that, when the vector time series $\Pi\mathbf{w}$ is partitioned as

$$\Pi\mathbf{w} = \begin{pmatrix}\mathbf{u}\\ \mathbf{y}\end{pmatrix}$$
$$\text{with } \mathbf{u}: \mathbf{R} \to \mathbf{k}^{q-p(\Sigma)}, \mathbf{y}: \mathbf{R} \to \mathbf{k}^{p(\Sigma)}$$

the relation induced by the behavior of Σ on $(\mathbf{u},\mathbf{y})$ will be a smooth I/O-system. We immediately conclude from this that $p(\Sigma)$ is the number of outputs in any (smooth) I/O representation $\Sigma_{I/O}$ of $\mathbf{L}^q$, while $m(\Sigma) := q - p(\Sigma)$ is the corresponding number of inputs.

For the problem at hand, i.e., computing and classifying all linear models of the polynomial-exponential set of data **D**, the following stronger result holds true.

Proposition 3.16: For any given choice of the input and output variables $\mathbf{u},\mathbf{y}$, there exist linear I/O unfalsified models of **D**.

Proof: Without loss of generality we may assume that the entries of **w** have been permuted so that $\mathbf{w} = \begin{pmatrix}\mathbf{u}\\ \mathbf{y}\end{pmatrix}$. Let Θ^* be partitioned accordingly, i.e., $\Theta^* = (\Theta_u\ \Theta_y)$. Due to the fact that Σ^* is an autonomous system, $\det\Theta^*$

$\neq 0$. It follows that each one of the two matrices above has full column rank, namely m and p, respectively. There exists therefore a submatrix of Θ_y of size p, denoted by $\mathbf{T}$, which is nonsingular. The corresponding rows of Θ_u define $\mathbf{Q}$, and together $(\mathbf{Q} \quad -\mathbf{T})$, define an I/O unfalsified model of $\mathbf{D}$. □

Corollary 3.17: For any given choice of the input and output variables, there also exist smooth I/O unfalsified models.

Proof: From (3.11) follows that without loss of generality Θ^* can be assumed to be in row reduced form. Thus, $[\Theta_u \quad \Theta_y]_{hr}$ is a constant square nonsingular matrix, and consequently $[\Theta_y]_{hr}$ has full column rank p; this assures the existence of a nonsingular submatrix of size p. The corresponding rows define $\mathbf{Q}$ and $\mathbf{T}$ satisfying property (3.15b), which implies the properness of Z. □

2) The Row-Degree Structure of Θ

Consider again (3.12). Writing these g equations explicitely yields

$$\begin{aligned} &\theta_1\left(\frac{d}{dt}\right)\mathbf{w} = 0 \\ &\theta_2\left(\frac{d}{dt}\right)\mathbf{w} = 0 \qquad \text{with } \Theta =: \begin{pmatrix}\theta_1\\ \theta_2\\ \vdots \\ \theta_g\end{pmatrix} \quad \theta_i \in \mathbf{k}^{1\times q}[s]. \\ &\vdots \\ &\theta_g\left(\frac{d}{dt}\right)\mathbf{w} = 0 \end{aligned} \tag{3.18}$$

Now, consider the ith equation: $\theta_i(d/dt)\mathbf{w} = 0$. Define the ith row degree $L_i := deg\theta_i$, and the *total row-degree* L to be

$$L := L_1 + L_2 + \cdots + L_g.$$

Among the minimal representations of a given $\Sigma \in \mathbf{L}^q$, there is obviously one whose total row degree is minimal. This minimum is a second integer invariant of Σ which will be of interest. It will be denoted by $n(\Sigma)$. Whence n is a mapping $n: \mathbf{L}^q \to \mathbf{Z}_+$.

It is possible to classify which minimal systems $\Sigma(\Theta)$ are such that their total row degree equals $n(\Sigma)$. Indeed, this is the case if, and only if, $\Theta \in \mathbf{k}^{g\times q}[s]$, is a *row reduced* polynomial matrix. It can be shown that $n(\Sigma)$ is the dimension of the state space in any minimal state space representation of Σ. This result is sometimes referred to as *Chrystal's Theorem*. For details, see Schumacher [23] and the references therein.

3) The Complexity

We are now in a position to define the complexity c of a dynamical system in $\mathbf{L}^q$.

$$c: \mathbf{L}^q \to \{0, 1, \cdots, q\} \times \mathbf{Z}_+$$
$$\text{where } c(\Sigma) := (m(\Sigma), n(\Sigma)). \tag{3.19}$$

The significance of this pair is twofold. Firstly, in terms of behavioral equations, $q - m(\Sigma)$ equals the minimal number of scalar differential equations needed to specify Σ, while $n(\Sigma)$ equals the minimal total row degree possible in these equations. Secondly, in terms of inputs and states, $m(\Sigma)$ equals the number of inputs in Σ, and $n(\Sigma)$ equals the number of internal states.

Now, endow the range of c (i.e., the complexity space) $\{0, 1, \cdots, q\} \times \mathbf{Z}_+$, with the *lexicographic*[9] ordering. This induces a pre-ordering on $\mathbf{L}^q$ by defining

$$(\Sigma_1 \leq \Sigma_2) :\Leftrightarrow (c(\Sigma_1) \leq c(\Sigma_2)).$$

In our modeling procedures, we will be paying special attention to models whose complexity is as small as possible in this sense.

Recall Definition 3.5 of an autonomous system. It is possible to prove (see Willems [26]) that the following conditions are equivalent: i) Σ is autonomous; ii) $p(\Sigma) = q$; iii) $m(\Sigma) = 0$; iv) $\mathbf{B}$ is finite dimensional. Thus, $\Sigma(\Theta)$ is autonomous iff $rank\Theta = q$, i.e., Σ is representable by a square nonsingular polynomial matrix Θ; in fact, the degree of det Θ equals the dimension of $\mathbf{B}$. Thus, according to (3.19), the complexity of Σ^* is the ordered pair:

$$c^* := c(\Sigma^*) = (0, \dim_k \mathbf{B}^*) = (0, \deg\det \Theta^*),$$

which implies that in the lexicographic ordering, the complexity of any autonomous system is always *less* than the complexity of any nonautonomous system.

E. Controllability

Among all linear models of $\mathbf{D}$, of particular interest are the *controllable* ones. A further important aspect at which the behavioral formalism departs from, and generalizes, the classical formalism is that related to controllability; controllability becomes namely an attribute of the *system* (i.e., of a collection of trajectories) as opposed to an attribute of a *system representation* (i.e., of equations generating these trajectories).

Roughly speaking, a system is *controllable* if its behavior has the property: whatever the *past* history (trajectory), it can always be *steered* to any desired *future* trajectory. More precisely, a dynamical system $\Sigma = (\mathbf{R}, \mathbf{k}^q, \mathbf{B}) \in \mathbf{L}^q$ is said to be *controllable* if for any $\mathbf{w}_1, \mathbf{w}_2 \in \mathbf{B}$, there exists *a* $t' > 0$ and a $\mathbf{w} \in \mathbf{B}$ such that

$$\mathbf{w}(t) = \begin{cases} \mathbf{w}_1(t) & \text{for } t < 0 \\ \mathbf{w}_2(t) & \text{for } t > t' \end{cases}.$$

In terms of AR behavioral equation representations, the system $\Sigma(\Theta)$ is controllable if, and only if, the rank of the (constant) matrix $\Theta(\lambda) \in \mathbf{C}^{g\times q}$ is constant and therefore equal to $p(\Sigma)$, for all $\lambda \in \mathbf{C}$, In particular, an I/O system [cf. (3.13a, b)] will be controllable if, and only if, the polynomial matrices $\mathbf{T}, \mathbf{Q}$ are left co-prime. The question of existence of controllable models of $\mathbf{D}$ is answered next.

Proposition 3.20: There exist nontrivial[10] controllable models of $\mathbf{D}$ if, and only if, Θ^* has at least one invariant factor equal to unity.

[9] The lexicographic ordering is a total ordering. It is defined as follows: given the vectors of n real numbers a, b we write $a \geq b$ if $a = b$ or if for some $j \in \underline{n}$, $a_i = b_i$, $i < j$, and $a_j > b_j$.

[10] The space $(\mathbf{k}^q)^{\mathbf{R}}$ of all time series is a trivial controllable model.

In Section VI it will be shown that this condition can be expressed directly in terms of properties of the data **D**; it will be shown namely, that the existence of controllable models is equivalent with the condition of having less that q linearly independent time series at each different frequency [see (6.7b)]. Thus, combining Proposition 3.16 and Corollary 3.17, with Proposition 3.20 we obtain the following lemma.

Lemma 3.21: The existence of controllable, I/O, smooth models of **D**. Assuming that there exist controllable models of **D**, there exist controllable, I/O, as well as controllable, I/O, smooth models of **D**. Clearly, minimal-complexity models with the above properties exist as well.

The class of controllable elements of $\mathbf{L}^q$ will be denoted by $\mathbf{L}^q_{contr}$. An MPUM in $\mathbf{L}^q_{contr}$ will *not* exist in general. This is due to the fact that controllability is not preserved under intersection of behaviors. For example, let $\Theta_i := (a_i \; b_i) \in \mathbf{k}^{1\times 2}[s]$, $i = 1, 2$, satisfy: a_i, b_i co-prime for $i = 1, 2$, and $\det \Theta_{12} \neq 0$, where $\Theta_{12} := \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}$. From (3.10b) follows that the intersection of the controllable systems with behaviors $\mathbf{B}(\Theta_1), \mathbf{B}(\Theta_2)$, is an autonomous system. Furthermore, if $\mathbf{L}^q_{contr}$ is nonempty, each $\Sigma \in L^q_{contr}$ with $p(\Sigma)$ fixed, is an undominated system, irrespective of its complexity. These facts are illustrated in the following

Example 3.22: Let $q = 2$, and consider the data set $\mathbf{D} := \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-t}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$, i.e., $\lambda_1 = -1$ and $\lambda_2 = 0$. It readily follows that $\Sigma(\Theta^*)$, where

$$\Theta^* := \begin{pmatrix} s+1 & -1 \\ 0 & s \end{pmatrix},$$

is the MPUM of **D** in $\mathbf{L}^2$. The family of systems $\Sigma(\Theta(\alpha))$, where

$$\Theta(\alpha) := (s+1 \quad \alpha s - 1), \qquad \alpha \in \mathbf{k}$$

satisfies $\Sigma(\Theta^*) \subset \Sigma(\Theta(\alpha))$ for all $\alpha \in \mathbf{k}$, i.e. by (3.11) $\Sigma(\Theta(\alpha))$ is a family of linear models of **D**. These systems are controllable for $\alpha \neq -1$. According to Section III-D-3) for these systems $m = n = 1$, i.e., their complexity $(1, 1)$ is minimal. Thus the family of systems just described, is the family of all minimal-complexity controllable models of **D**.

Let $\Theta_i := (s+1 \;\; \alpha_i s - 1)$ $\alpha_i \neq -1$, $i = 1, 2$, $\alpha_1 \neq \alpha_2$. It follows that the greatest common right divisor of these two matrices is Θ^*. Thus, according to (3.10b) the intersection of the two families $\mathbf{B}(\Theta_i)$, $i = 1, 2$, of controllable models of **D**, is equal to the MPUM of **D**. Clearly, there is *no* controllable MPUM in this case. The family of all controllable models of **D** can be obtained using the above formula, where the parameter α is any polynomial satisfying $\alpha(-1) \neq -1$. It is readily checked that each one of these models is undominated. □

In the sequel we will make use of the following classes of linear models of **D**, parametrized in terms of vector parameters α, β, γ of appropriate dimension:

$\Sigma^*_{contr}(\alpha)$: minimal-complexity controllable models;

$$c^*_{contr} := c(\Sigma^*_{contr}(\alpha)) \quad (3.23a)$$

$\Sigma^*_{I/O}(\beta)$: minimal-complexity controllable I/O models;

$$c^*_{I/O} := c\left(\Sigma^*_{I/O}(\beta)\right) \quad (3.23b)$$

$\Sigma^*_{pr}(\gamma)$: minimal-complexity controllable smooth I/O

$$\text{models}; \; c^*_{pr} := c\left(\Sigma^*_{pr}(\gamma)\right). \quad (3.23c)$$

In fact, these parametrizations can be taken as *affine* in the parameters (see Theorem 9.9 and Remark 9.20a). It should be remarked that the members of the above three families are undominated models, and therefore these families can be compared with one-another only by means of their complexities.

Proposition 3.24: The complexities of the MPUM and of the above three families of models of **D**, satisfy the following inequalities:

$$c^* \leq c_{contr} \leq^* c^*_{I/O} \leq c^*_{pr}$$

Example A of Section XI, illustrates the facts related to Lemma 3.21. It also demonstrates that the inequalities above can be strict.

1) The Transfer Function

As already mentioned, given the I/O system (3.13a, b), the matrix of rational functions Z defined by (3.15a) is called the *transfer function*. Two I/O systems may have the same transfer function without having the same behavior, since the common factors of **T** and **Q** might be different in the two cases. However, it can be shown that two controllable I/O systems have the same transfer function if, and only if, they have the same behavior. Consequently, the transfer function parametrizes in a unique way the controllable I/O systems (in Willems [29] this is referred to as a trim parametrization).

2) Moving Average (MA) Behavioral Equations

Following the general theory of the behavioral framework (see above mentioned references) it follows that any $\Sigma_{contr} = (\mathbf{T}, \mathbf{k}^q, \mathbf{B}_{contr}) \in \mathbf{L}^q_{contr}$, can be represented in terms of *moving average* (abbreviated MA) behavioral equations. Recall (3.13a, b). Let $\hat{\mathbf{Q}} \in \mathbf{k}^{p\times m}[s]$ and $\hat{\mathbf{T}} \in \mathbf{k}^{m\times m}[s]$, with $\det \hat{\mathbf{T}} \neq 0$, be right co-prime polynomial matrices satisfying:

$$\mathbf{Q}\hat{\mathbf{T}} = \mathbf{T}\hat{\mathbf{Q}}.$$

It follows that the columns of $\begin{pmatrix} \hat{\mathbf{T}} \\ \hat{\mathbf{Q}} \end{pmatrix}$ define a basis for the kernel of $\Theta = (\mathbf{Q} \quad -\mathbf{T})$. Thus, from (3.13a) $(\mathbf{u}, \mathbf{y}) \in \mathbf{B}_{I/O}$, defined by (3.14) if, and only if, there exists a sufficiently smooth $\mathbf{a}: \mathbf{T} \to \mathbf{k}^m$, such that $\mathbf{u} = \hat{\mathbf{T}}(d/dt)\mathbf{a}$ and $\mathbf{y} = \hat{\mathbf{Q}}(d/dt)\mathbf{a}$. This means that the behavior $\mathbf{B}_{contr}$ can be expressed as the *image* of an appropriately defined polynomial operator Ψ_{contr} in the derivative operator (d/dt):

$$\mathbf{B}_{\text{contr}} = im\Psi_{\text{contr}}\left(\frac{d}{dt}\right)$$

$$\text{where } \Psi_{\text{contr}}\left(\frac{d}{dt}\right) := \begin{pmatrix} \hat{\mathbf{T}}\left(\frac{d}{dt}\right) \\ \hat{\mathbf{Q}}\left(\frac{d}{dt}\right) \end{pmatrix}.^{11} \tag{3.25}$$

Thus, $\mathbf{w} = \Psi_{contr}(d/dt)\mathbf{a}$ for $\mathbf{a}$ as above. The entries of $\mathbf{a}$ are the *auxiliary* variables mentioned at the beginning of Section III-C. The MA representation thus provides a direct[12] characterization of the controllable behaviors. Notice that the transfer function (3.15a) can be written as a left or right matrix fraction:

$$Z = \mathbf{T}^{-1}\mathbf{Q} = \hat{\mathbf{Q}}\hat{\mathbf{T}}^{-1}.$$

This means that for controllable systems, AR representations correspond to left co-prime matrix fractions of Z, while (observable) *MA* representations correspond to (right) co-prime matrix fractions of Z.

F. The Main Problem (Precise Statement)

Find all linear models of the data set $\mathbf{D}$ defined by (1.1)–(1.3). This means, find all $\Sigma = (\mathbf{R}, \mathbf{k}^q, \mathbf{B}) \in \mathbf{L}^q$ such that $\mathbf{D} \subset \mathbf{B}$. According to Section III-B, this amounts to computing the MPUM $\mathbf{B}^*$ of $\mathbf{D}$. Subsequently, find behavioral equation representations of $\mathbf{B}^*$ [cf. (3.6a), (3.7a)]. Special attention is to be paid to questions of *recursiveness* in the number n of time series. Furthermore, the recursive update of both the MPUM $\mathbf{B}^*$ and of its behavioral equation representations is to be worked out. Important properties of the models to keep tract of, are *model complexity* [cf. (3.19)] and *model controllability* (cf. Section III-V). In particular, the recursive update of minimal-complexity controllable models (C-MCUM's) is to be described. Other properties to be considered are *a priori* imposed input–output structure [cf. (3.14)] and *model smoothness* (cf. Section III-D).

IV. The MPUM (Most Powerful Unfalsified Model) Σ^* of $\mathbf{D}$

Consider the data set $\mathbf{D}$ defined in (1.1)–(1.3). From the given time series $\mathbf{w}_i$, we generate a set of related time series by differentiating an appropriate number of times:

$$\mathbf{w}_{i,r-1} := \left[\frac{d}{dt} - \lambda_i\right]^{r-1} \mathbf{w}_i = \mathbf{p}_{i,r-1}\exp_{\lambda_i}, \qquad r \in \underline{\kappa}_i, \tag{4.1}$$

where

$$\mathbf{p}_{i,r-1}(t) := \frac{d^{r-1}}{dt^{r-1}}\mathbf{p}_i(t) = \sum_{j=r}^{\kappa_i} p_{i,\kappa_i-j}\frac{t^{j-r}}{(j-r)!}.$$

Thus, $\mathbf{w}_{i,0} = \mathbf{w}_i$. Using the above set of time series, we define the following linear spaces:

$$\mathbf{B}_i^* := \text{span}_k\{\mathbf{w}_{i,r-1}: r \in \underline{\kappa}_i\}, \qquad i \in \underline{n} \text{ and } \mathbf{B}^* := \sum_{i\in\underline{n}} \mathbf{B}_i^*. \tag{4.2}$$

For subsequent use, to $\mathbf{B}_i^*, \mathbf{B}^*$ we associate the polynomials

$$\pi_i := (s - \lambda_i)^{\kappa_i} \quad \text{and} \quad \pi := \prod_{i\in\underline{n}} \pi_i, \tag{4.3}$$

respectively, which will be referred to as the *characteristic polynomials*. The main result of this section, which solves our modeling problem on the level of behaviors, is the following theorem.

Theorem 4.4: The system $\Sigma^* = (\mathbf{R}, \mathbf{k}^q, \mathbf{B}^*)$, where $\mathbf{B}^*$ is defined by (4.2), is the MPUM (most powerful unfalsified model) of the data set $\mathbf{D}$, defined by (1.1)–(1.3) in the class of linear systems $\mathbf{L}^q$.

The characterization of this and all other models of $\mathbf{D}$ in $\mathbf{L}^q$ in terms of equations (which we have called *behavioral equations*) is given in the following three sections. It should be remarked that the results of these sections have connections with the theory of *polynomial models* and *shift realizations* introduced by Fuhrmann [14], [16]. These connections will be pursued elsewhere.

Proof: By construction, $\mathbf{B}_i^*$ is a linear space. A straightforward computation yields the identity

$$\sigma^T\mathbf{w}_i = \left(\mathbf{w}_{i,0} + \frac{T}{1!}\mathbf{w}_{i,1} + \frac{T^2}{2!}\mathbf{w}_{i,2} + \cdots + \frac{T^{\kappa_i-1}}{(\kappa_i-1)!}\mathbf{w}_{i,\kappa_i-1}\right)\exp_{\lambda_i}(T),$$

where σ denotes the backward shift, and the time series $\mathbf{w}_{i,j}$ are defined in (4.1). Clearly

$$\text{span}_{\mathbf{k}}\{\sigma^T\mathbf{w}_i: T \in \mathbf{R}\} = \text{span}_{\mathbf{k}}\{\mathbf{w}_{i,0}, \mathbf{w}_{i,1}, \cdots, \mathbf{w}_{i,\kappa_i-1}\} = \mathbf{B}_i^*.$$

Therefore, $\mathbf{B}_i^*$ is the smallest linear, shift invariant space which contains the time series $\mathbf{w}_i$. Moreover, it is differential. Consequently, $\mathbf{B}^*$ is the smallest linear, shift invariant, instantaneous specified subspace of $(\mathbf{k}^q)^{\mathbf{R}}$ which contains $\mathbf{D}$. Therefore Σ^* is indeed, the MPUM of $\mathbf{D}$ in $\mathbf{L}^q$. □

From (4.1), (4.2) follows that $\dim_k \mathbf{B}_i^* = \kappa_i$. Recall (1.4). From (4.2):

$$\dim_k \mathbf{B}^* =: N^* \le N = \kappa_1 + \cdots + \kappa_n. \tag{4.5}$$

Clearly, equality holds if, and only if, the sum in (4.2) is a direct sum, in other words if, and only if, $\mathbf{B}_i^* \cap \mathbf{B}_j^* = 0$, $i \ne j$. In Proposition 4.7 below we shall express this condition directly as a condition in terms of the time series (1.1); notice that in (1.1) the case $\lambda_i = \lambda_j$, $i \ne j$, is not excluded. To achieve this goal, we need to partition $\mathbf{D}$

[11] $\Psi_{contr}(d/dt)$ is interpreted as a map from $C^\infty(\mathbf{R}, \mathbf{k}^q)$ to $C^\infty(\mathbf{R}, \mathbf{k}^p)$.
[12] *Direct/Annihilating* characterization of a behavior: characterization as the *image/kernel* of an operator.

according to the different frequencies of the time series $\mathbf{w}_i$. Among λ_i, $i \in \underline{n}$, let there be $l \leq n$ different frequencies, denoted by:

$$\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_l, \hat{\lambda}_i \neq \hat{\lambda}_j, \qquad i \neq j, \tag{4.6a}$$

with m_i time series at each one of these frequencies $\hat{\lambda}_i$, $i \in \underline{l}$. The data set $\mathbf{D}$ is now partitioned accordingly:

$$\mathbf{D} = \mathbf{D}_1 \cup \mathbf{D}_2 \cup \cdots \cup \mathbf{D}_l \tag{4.6b}$$

where

$$\mathbf{D}_i := \mathbf{P}_i \exp_{\hat{\lambda}_i}, \quad \mathbf{P}_i \in \mathbf{k}^{q \times m_i}[\cdot],$$

$$m_1 + m_2 + \cdots + m_l = n; \tag{4.6c}$$

thus, $\mathbf{P}_i$ is a polynomial matrix, whose columns consists of all those polynomial vectors $\mathbf{p}_j$ in (1.1), which correspond to time series having the same frequency. Mainly, for later use, we also introduce the positive integer

$$m := \max_{i \in \underline{l}} \{m_i\} \geq 1. \tag{4.6d}$$

We are now ready to answer the question posed earlier.

Proposition 4.7: With the notation introduced above, the following statements are equivalent.

i) $N^* = N$.

ii) $\mathbf{B}_i^* \cap \mathbf{B}_j^* = 0$ for $i \neq j$.

iii) Each one of the polynomial matrices $\mathbf{P}_i$, $i \in \underline{l}$, is column reduced.

For a definition of column reducedness, see Kailath [19]. Next we will show that any one of the data sets $\mathbf{D}_i$ defined in (4.6c) can be pre-processed so that condition iii) of the above proposition is satisfied.

Proposition 4.8: Given the time series $\mathbf{w}_i = \mathbf{p}_i \exp_\lambda$, $i \in \underline{m}$, and the associated linear spaces $\mathbf{B}_i^*$, there exist time series $\hat{\mathbf{w}}_i = \hat{\mathbf{p}}_i \exp_\lambda$, $\hat{\mathbf{p}}_i \in \mathbf{k}^q[\cdot]$, such that the associated linear spaces $\hat{\mathbf{B}}_i^*$, satisfy:

$$\sum_{i \in \underline{m}} \mathbf{B}_i^* = \sum_{i \in \underline{m}} \hat{\mathbf{B}}_i^* \quad \text{and} \quad \hat{\mathbf{B}}_i^* \cap \hat{\mathbf{B}}_j^* = 0 \qquad \text{for } i \neq j.$$

Proof: For simplicity, suppose that $m = 2$, $\deg \mathbf{p}_1 = \kappa_1 \geq \kappa_2 = \deg \mathbf{p}_2$, and let Proposition 4.7-iii) not be satisfied. This means that $p_{1,0} = \alpha p_{2,0}$, for some $\alpha \in \mathbf{k}$, that is $\mathbf{B}_1^* \cap \mathbf{B}_2^* \neq 0$. Define

$$\mathbf{w}_2^{\text{new}} := \mathbf{p}_2^{\text{new}} \exp_\lambda \quad \text{and} \quad \mathbf{p}_2^{\text{new}} := \mathbf{p}_{1, \kappa_1 - \kappa_2} - \alpha \mathbf{p}_2^{\text{old}},$$

where, as in (4.1), $\mathbf{p}_{1,\kappa}$ denotes the κth derivative of $\mathbf{p}_1$ with respect to time. Clearly the degree κ_2^{new} of $\mathbf{p}_2^{new}$ is strictly less than $\kappa_2^{old} := \kappa_2$. Moreover,

$$\mathbf{B}_1^* + \mathbf{B}_2^* = \mathbf{B}_1^* + \mathbf{B}_2^{*,\text{new}}.$$

If $\mathbf{B}_1^* \cap \mathbf{B}_2^{*,new} = 0$, take $\hat{\mathbf{p}}_2 := \mathbf{p}_2^{new}$. Otherwise, repeat the above procedure with $\mathbf{w}_2$ replaced by $\mathbf{w}_2^{new}$. This algorithm can readily be generalized to more than two time series. □

Remark 4.9: To summarize, the dimension of $\mathbf{B}^*$ is bounded from above by the sum of the degrees of the underlying polynomial vectors $\mathbf{p}_i$. In order for equality to hold, for each group of time series having the same frequency, the column reducedness condition iii) of Proposition 4.7 must hold. The algorithm for obtaining such a column reduced matrix, which is described above in the *time* domain, does not correspond to the usual postmultiplication of $\mathbf{P}_i$ by a unimodular matrix. It is easy to check however, that this property does hold if these operations are performed in the s-domain. □

V. SV (State Variable) Behavioral Equations for Σ^*

The MPUM Σ^* of $\mathbf{D}$ in the class of linear systems $\mathbf{L}^q$, computed in the previous section has *finite-dimensional* behavior [cf. (4.5)]. As already mentioned in Section III-D-3), the latter property implies that Σ^* is *autonomous*. It follows from the general theory developed by Willems [27], [29], that the MPUM can be represented in terms of SV (state variable) equations of the form (3.6a). We start by writing state variable behavioral equations for $\mathbf{B}_i^*$ based on (1.1), (1.2), (4.1). Consider the equations

$$\frac{d}{dt}\mathbf{x}_i = F_i \mathbf{x}_i, \quad \mathbf{w} = H_i \mathbf{x}_i$$

$$\text{where } F_i := \lambda_i I + J,\ H_i := (p_{i,0} \cdots p_{i,\kappa_i - 1}); \tag{5.1}$$

I is the identity matrix of size κ_i, J is the nilpotent matrix of the same size with 1s on the superdiagonal and 0s elsewhere, and $\mathbf{x}_i \in (\mathbf{k}^{\kappa_i})^{\mathbf{R}}$. Notice that the characteristic polynomial π_i of $\mathbf{B}_i^*$ introduced in the previous section, is equal to the characteristic polynomial of F_i defined above:

$$\pi_i = \det(sI - F_i). \tag{5.2}$$

Furthermore, H_i, F_i is an *observable* pair because, by (1.2), $p_{i,0} \neq 0$. The following holds true.

Lemma 5.3: Equations (5.1) constitute a minimal-complexity SV equation representation of $\mathbf{B}_i^*$.

Proof: The expression

$$\exp(F_i t) = \exp(\lambda_i t)\left(I + Jt + \cdots + J^{\kappa_i - 1}\frac{t^{\kappa_i - 1}}{(\kappa_i - 1)!}\right)$$

can be readily computed from (5.1), Therefore the pair

$$\hat{\mathbf{x}} := \exp(F_i t)v, \quad \hat{\mathbf{w}} := H_i \hat{\mathbf{x}},$$

satisfies the equations in (5.1), for all $v \in \mathbf{k}^{\kappa_i}$. It follows that if we let $v := \zeta_r$, the rth unit vector, $\hat{\mathbf{w}} = \mathbf{w}_{i,r-1}$, $r \in \underline{\kappa_i}$. Therefore by linearity, every trajectory contained in $\mathbf{B}_i^*$ is generated by (5.1) and conversely, every trajectory generated by (5.1) belongs to $\mathbf{B}_i^*$. Finally, minimality is a consequence of the observability of the pair H_i, F_i. □

A state variable representation of the MPUM Σ^* for $\mathbf{B}^*$ can now be put together as a direct sum of the representation of the MPUMs of the time series computed above. let

$$H := (H_1 \cdots H_i \cdots H_n) \in \mathbf{k}^{q \times N},$$

$$F := \operatorname{diag}(F_1, \cdots, F_i, \cdots, F_n) \in \mathbf{k}^{N \times N}. \tag{5.4}$$

The following result holds.

Theorem 5.5: Equations (3.6a), with H, F defined by (5.4), constitute a SV equation representation of the MPUM Σ^* of $\mathbf{D}$, constructed in Theorem 4.4.

Analogously to (5.2), from (4.3) and (5.4) follows that $\pi = \det(sI - F)$. Moreover, from Proposition 4.7 follows the corollary below.

Corollary 5.6: The SV representation of the theorem has minimal complexity if, and only if, the equality $N^* = N$ holds if, and only if, the pair H, F is observable. In this case with the notation introduced in (3.6b), $\mathbf{B}^* = \mathbf{B}(H, F)$.

VI. AR (Autoregressive) Behavioral Equations for Σ^*

The SV behavioral equations discussed in the previous section can be re-written in the form:

$$\Omega\left(\frac{d}{dt}\right)\begin{pmatrix}\mathbf{x}\\ \mathbf{w}\end{pmatrix} = 0$$

where $\Omega := \begin{pmatrix} sI - F & 0 \\ H & -I \end{pmatrix} \in \mathbf{k}^{(q+N)\times(q+N)}[s]$ and

$$\mathbf{x} \in (\mathbf{k}^N)^{\mathbf{R}}, \mathbf{w} \in (\mathbf{k}^q)^{\mathbf{R}}. \quad (6.1)$$

In the above expressions s is to be interpreted as the derivative operator. Let the polynomial matrices

$$\Xi \in \mathbf{k}^{q\times N}[s], \quad \Theta^* \in \mathbf{k}^{q\times q}[s], \qquad (6.2)$$

be *left co-prime*, and satisfy

$$\Xi(sI - F) = \Theta^* H. \qquad (6.3)$$

Recall notation 3.9. The following crucial result holds.

Theorem 6.4: Θ^* is an autoregressive equation representation of the MPUM Σ^* of $\mathbf{D}$. In other words $\Sigma^* = \Sigma(\Theta^*)$, i.e.,

$$\Theta^*\left(\frac{d}{dt}\right)\mathbf{w} = 0 \qquad \text{for all } \mathbf{w} \in \mathbf{B}^*,$$

where Θ^* is defined by (6.2), (6.3).

Proof: We will assume that $\mathbf{B}^*$ is parametrized in terms of the observable pair H, F, i.e., by corollary (5.6), $\mathbf{B}^* = \mathbf{B}(H, F)$. Thus, we need to show that $\mathbf{B}(\Theta^*) = \mathbf{B}(H, F)$.

Let $\mathbf{w} \in \mathbf{B}(H, F)$. It follows that $\mathbf{w} = H\mathbf{x}$ and $s\mathbf{x} = F\mathbf{x}$ for an appropriate $\mathbf{x}$. Upon multiplying (6.1) on the left by $(\Xi \quad -\Theta^*)$, we obtain $\Theta^*\mathbf{w} = 0$, which shows that $\mathbf{w} \in \mathbf{B}(\Theta^*)$.

Conversely, let $\mathbf{w} \in \mathbf{B}(\Theta^*)$, i.e., $\Theta^*\mathbf{w} = 0$. We will show that there exists a state trajectory $\mathbf{x}$, such that $\mathbf{w} = H\mathbf{x}$ and $s\mathbf{x} = F\mathbf{x}$. For the proof of this inclusion, we need to use the following fact which can be found, e.g., in Kailath [19, ch. 6.4]. Given the observable pair H, F and the left co-prime matrices Ξ, Θ^* as defined above, there exist polynomial matrices A, B of appropriate size, such that the following properties hold:

$$K := \begin{pmatrix} A & B \\ \Xi & -\Theta^* \end{pmatrix} \in \mathbf{k}^{(q+N)\times(q+N)}[s]$$

$$\text{with } \det K = 1 \text{ and } K\begin{pmatrix} sI - F \\ H \end{pmatrix} = \begin{pmatrix} I \\ 0 \end{pmatrix}.$$

Define $\mathbf{x} := B\mathbf{w}$. Then

$$K\Omega\begin{pmatrix}\mathbf{x}\\ \mathbf{w}\end{pmatrix} = \begin{pmatrix} I & -B \\ 0 & \Theta^* \end{pmatrix}\begin{pmatrix} B\mathbf{w} \\ \mathbf{w}\end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ \Theta^*\mathbf{w}\end{pmatrix} = 0 \Rightarrow \begin{pmatrix}\mathbf{x}\\ \mathbf{w}\end{pmatrix} \in \ker K\Omega.$$

Since K is unimodular, by Proposition 3.10-a3) $\ker K\Omega = \ker \Omega$. Thus, $\mathbf{w} \in \mathbf{B}(H, F)$, and the proof is complete. □

From (3.11) furthermore, we obtain a result which will be used in Section VIII, on recursiveness.

Theorem 6.5: Let $\Sigma(\hat{\Theta})$ be a model of $\mathbf{D}$. $\Sigma(\hat{\Theta}) = \Sigma(\Theta^*)$, if, and only if, the invariant factors of $\hat{\Theta}$ and Θ^* are the same, say $\Delta_1|\Delta_2|\cdots|\Delta_q$.

In the sequel we will compute these invariant factors in terms of the characteristic polynomials π_i, $i \in \underline{n}$. For this purpose we will assume without loss of generality, that the data set has been pre-processed so that Conditions 4.7-iii) are satisfied, i.e., so that the SV representation has minimal complexity. According to Corollary 5.6 this is equivalent to the observability of the pair H, F (i.e., the right co-primeness of the polynomial matrices $H, sI\text{-}F$); since Ξ, Θ^* are left co-prime, it follows from (6.3) (see e.g., Fuhrmann [14] or Kailath [19, Section 6.4.2]) that the nontrivial invariant factors of $(sI - F)$ and $\Theta^*(s)$ are the same. Recall the Definitions 4.3, 5.2 of the characteristic polynomials. Using the abbreviation *gcd* for *greatest common divisor*, we define the following n polynomials:

$$\delta_1 := gcd(\pi_i, i \in \underline{n}),$$

$$\vdots$$

$$\delta_j := gcd(\pi_{i_1}\pi_{i_2}\cdots\pi_{i_j}, i_\kappa \in \underline{n}, \kappa \in \underline{j}),$$

$$\vdots$$

$$\delta_n := \pi_1 \cdots \pi_j \cdots \pi_n = \pi. \qquad (6.6)$$

Consequently, with F given by (5.4), it follows that the invariant factors of Θ^* are

$$\Delta_q := \frac{\delta_n}{\delta_{n-1}}, \cdots, \Delta_{q-j+1} := \frac{\delta_{n-j-1}}{\delta_{n-j}}, \cdots, \Delta_{q-n+1} := \delta_1; \qquad (6.7a)$$

if $n < q$, there are $q - n$ additional unity invariant factors; if $n \geq q$, at most the first q of the above invariant factors will be different from unity. Clearly, $\det \Theta^* = \Delta_1 \cdots \Delta_n = \pi \neq 0$. Actually, with m defined by (4.6d), there are exactly m[13] nontrivial (i.e., nonunity) invariant factors among the Δ_i, namely:

$$\Delta_q, \Delta_{q-1}, \cdots, \Delta_{q-m-2}, \Delta_{q-m+1}, 1, \cdots, 1. \qquad (6.7b)$$

Remark 6.8: *i*) If there is only one polynomial-exponential time series at each frequency, i.e., in (1.1) $\lambda_i \neq \lambda_j$, $i \neq j$, any two π_i, $i \in \underline{n}$, are co-prime and hence, in this case, the invariant factors of Θ^* are $\pi, 1, \cdots, 1$. ii) Other-

[13] According to Remark 6.10, it can be assumed w.l.o.g. that $m < q$.

wise, let the column degrees of the polynomial matrix $\mathbf{P}_i$ in (4.6c) be: $\hat{\kappa}_1 \leq \cdots \leq \hat{\kappa}_{m_i}$. It follows that

$$\left(s-\hat{\lambda}_i\right)^{\hat{\kappa}_{m_i}}|\Delta_q,\cdots,\left(s-\hat{\lambda}_i\right)^{\hat{\kappa}_1}|\Delta_{q-m_i+1}.$$

Using this fact, given the data set $\mathbf{D}$ and its partitioning defined in (4.6a–c), the invariant factors of the MPUM can be written down by inspection. □

Remarks 6.9: a) The AR equation representation Θ^* of Σ^* derived above, and the unimodular matrix $\mathbf{W}$ which is the main tool in Antoulas [1], are closely related. Without loss of generality, assume that Θ^* is row reduced with row indexes $\upsilon_1,\cdots,\upsilon_q$. Let

$$\tilde{\Theta}(s^{-1}) := \operatorname{diag}(s^{-\upsilon_1},\cdots,s^{-\upsilon_q})\Theta^*(s).$$

It readily follows that $\tilde{\Theta}$ is a polynomial unimodular matrix in the variable s^{-1}. Thus

$$\mathbf{W}(s) = \tilde{\Theta}(s).$$

This shows that up to a change of variables, the $\mathbf{W}$ matrix introduced in Antoulas [1, (4.4a)] is indeed an AR equation representation of the MPUM of the corresponding modeling problem, i.e., the realization problem (see also Problem $\mathbf{D}$ of Section II).

b) Theorem 6.5 gives a condition for the minimality of the autonomous models simply in terms of their invariant factors. The corresponding result in the framework of [1], is given in Antoulas [9]. □

Remark 6.10: Assume that the data contains the following set of time series

$$\mathbf{D}_\lambda := \{p_1\exp_\lambda, p_2\exp_\lambda,\cdots,p_\kappa\exp_\lambda;\ p_i \in \mathbf{k}^q\},$$

where $\operatorname{rank}[p_1\ p_2\ \cdots\ p_\kappa] = q$. This implies that

$$\zeta_i \exp_\lambda \in \mathbf{B}_\lambda, \qquad \text{for all } i \in \underline{q},$$

where ζ_i denotes the ith unit q-vector. It follows that the MPUM of $\mathbf{D}_\lambda$ is $\Sigma_\lambda = (\mathbf{R}, \mathbf{k}^q, \mathbf{B}_\lambda)$, where

$$\mathbf{B}_\lambda = \exp_\lambda \cdot \mathbf{k}^q.$$

Therefore, using Notation 3.9, $\Sigma_\lambda = \Sigma(\Theta_\lambda)$, where

$$\Theta_\lambda = (s-\lambda)I_q.$$

Since Θ_λ commutes with any matrix, it follows that the AR equation Θ of the MPUM of $\mathbf{D}$ containing $\mathbf{D}_\lambda$ will be of the form

$$\Theta^* = \Theta_\lambda\overline{\Theta} = (s-\lambda)\overline{\Theta},$$

for an appropriate $\overline{\Theta}$. Therefore, whenever $\mathbf{D}_\lambda \subset \mathbf{D}$ we need only model the time series contained in $\overline{\mathbf{D}} := \mathbf{D} - \mathbf{D}_\lambda$, subsequently making use of the above formula.

Notice also that this case represents a sharp departure from the usual interpolation problem which can be associated with the modeling of time series (see Problem C in Section II). The formula for Θ^* given above, implies namely, that the frequency λ is *both* a pole and a zero of any resulting I/O system; cancellation is *not* allowed. This means in turn, that in such a case, by Proposition 3.20, there is no nontrivial *controllable* model of $\mathbf{D}$. □

Remark 6.11: One way of computing Θ^* is by determining the linear dependences of the rows of the observability matrix of the pair H, F defined in (5.4). For details, see e.g., Antoulas, Ball, Kang, and Willems [8, Section 8]. □

VII. AR AND MA BEHAVIORAL EQUATIONS FOR CONTROLLABLE MODELS

After the SV and AR equation representations of Σ^*, we turn our attention to behavioral equation representations of controllable models of $\mathbf{D}$. We will describe both AR and MA, i.e., *moving average*, ones (see Section III-E-2)).

In general, an AR equation representation of the minimal-complexity models $\Sigma^*_{contr} \in \mathbf{L}^q_{contr}$ of $\mathbf{D}$, can be computed from Θ^* as follows. Recall from Section III-E that a system is controllable if, and only if, any AR equation representing its behavior has constant rank for all complex frequencies. Consider the Smith form decomposition of Θ^* (for a definition see, e.g., Kailath [19, p. 390]):

$$\Theta^* = U\Delta V,$$

where U, V are units in $\mathbf{k}^{q\times q}[s]$, and Δ is a diagonal matrix, composed of the invariant factors Δ_i, defined by (6.7a, b). Notice that in this decomposition, the invariant factors are unique, while U, V are *not* unique. Recall (4.6d) and define p so that

$$p + m = q;$$

according to (6.7b) the integer p is the number of unity invariant factors of Θ^*. Thus, we can write

$$\Delta = \operatorname{diag}\left(\tilde{\Delta}, I_p\right), \quad \tilde{\Delta} := \operatorname{diag}(\Delta_q, \Delta_{q-1},\cdots,\Delta_{q-m+1})$$

where according to (6.7b) all diagonal elements of $\tilde{\Delta}$ are nonunity polynomials. It follows that

$$\Sigma^*_{\text{contr}} = \Sigma(\Theta^*_{\text{contr}}) \qquad \text{where } \Theta^*_{\text{contr}} := (0\quad I_p)V^{14} \quad (7.1)$$

is a minimal-complexity controllable model. Actually, an AR equation representation of the other minimal-complexity controllable models of $\mathbf{D}$ can be obtained by means of the above formula, by letting V vary over all allowable units in the Smith form decomposition. As mentioned in Section III-E, it will be shown in Section IX-A, that the family of least-complexity controllable models can be parametrized in an affine way. Now following Section III-E-2) the corresponding MA behavioral equation representation of these systems is given by:

$$\mathbf{w} = \Psi^*_{\text{contr}}\left(\frac{d}{dt}\right)\mathbf{a} \qquad \text{where } \Psi^*_{\text{contr}} := V^{-1}\begin{bmatrix} I_m \\ 0 \end{bmatrix}, \quad (7.2)$$

the entries of $\mathbf{a}$ in the above equations being the so-called *auxiliary* variables.

[14] Recall that the superscript * denotes least-complexity-related quantities.

VIII. Recursive Update of the MPUM and of the Associated SV and AR Equations

Let the data $\mathbf{D}$ be updated by conducting a new experiment:

$$\hat{\mathbf{D}} := \mathbf{D} \cup \mathbf{D}_{n+1}{}^{15} \qquad \text{with } \mathbf{D}_{n+1} := \{\mathbf{w}_{n+1} := \mathbf{p}_{n+1}\exp_{\lambda_{n+1}}\}.$$

There are two basic cases to consider. *First* $\mathbf{p}_{n+1}$ is constant, i.e., using Notation 1.2

$$\mathbf{p}_{n+1} =: p_{n+1,0} \in \mathbf{k}^q, \qquad p_{n+1,0} \neq 0. \tag{8.1}$$

Second, $\lambda_{n+1} = \lambda_h$, for some $h \in \underline{n}$, and in addition $\mathbf{p}_{n+1}$ is a polynomial such that

$$\frac{d\mathbf{p}_{n+1}}{dt} = \mathbf{p}_h,$$

where $\mathbf{p}_h$ is defined in (1.2). Integrating the above relationship we get:

$$\mathbf{p}_{n+1} := p_{h,\kappa_h} + \sum_{j=1}^{\kappa_h} p_{h,\kappa_h - j}\frac{t^j}{j!}. \tag{8.2}$$

Thus, in the former case a pure exponential time series is provided at a frequency which might or might not be different from the previous ones. In the latter case the new datum corresponds to adding a new term to an already existing time series. To be more specific, the existing hth time series having frequency λ_h, is updated by integrating the polynomial $\mathbf{p}_h$; the new constant entering is p_{h,κ_h}.

Let $\hat{\Sigma}^* := (\mathbf{R}, \mathbf{k}^q, \hat{\mathbf{B}}^*)$, denote the MPUM of $\hat{\mathbf{D}}$ in $\mathbf{L}^q$. It is obtained from Σ^*, defined in Theorem 4.4, by updating $\mathbf{B}^*$ as follows:

$$\hat{\mathbf{B}}^* := \mathbf{B}^* + \mathbf{B}_{n+1} \qquad \text{where } \mathbf{B}_{n+1} := \text{span}_{\mathbf{k}}\{\mathbf{w}_{n+1}\}. \tag{8.3}$$

Moreover, the characteristic polynomial of $\hat{\mathbf{B}}^*$, defined in Section IV, is $\hat{\pi} := (s - \lambda_{n+1})\pi$, where π is the characteristic polynomial of $\mathbf{B}^*$. It follows that in the above two cases the MPUM is updated by adding a *one-dimensional* subspace of $(\mathbf{k}^q)^{\mathbf{R}}$. Other cases can be reduced to either one of the above two [cf. Remark 8.22-a)].

The problem now is to update the SV and AR equation representation of $\mathbf{B}^*$ given in Theorems 5.5 and 6.4, in order to obtain the corresponding representations of $\hat{\mathbf{B}}^*$.

The updated SV equation representation is determined as follows. Let

$$\hat{H} \in \mathbf{k}^{q \times (N+1)}, \quad \hat{F} \in \mathbf{k}^{(N+1)\times(N+1)}.$$

If (8.1) is satisfied, Formulae 5.4 are updated by attaching a new block:

$$\hat{H} := (H \quad p_{n+1,0}), \quad \hat{F} := \text{diag}(F, \lambda_{n+1}). \tag{8.4}$$

If (8.2) is satisfied, the hth block of H, F is updated:

$$\hat{H} := \left(H_1 \cdots \hat{H}_h \cdots H_n\right)$$

$$\text{where } \hat{H}_h := (H_h \quad p_{h,\kappa_h}) \in \mathbf{k}^{q\times(\kappa_h+1)}, \tag{8.5a}$$

$$\hat{F} := \text{diag}\left(F_1, \cdots, \hat{F}_h, \cdots, F_n\right)$$

$$\text{where } \hat{F}_h := \lambda_h I + J \in \mathbf{k}^{(\kappa_h+1)\times(\kappa_h+1)}. \tag{8.5b}$$

It is now a simple matter, following Theorem 5.5, to show that equations (3.6a), with $\hat{H}, \hat{F}$ as defined above, constitute indeed an SV equation representation of $\hat{\mathbf{B}}^*$.

The next step is to update the AR equation representation Θ^*, which we will denote by $\hat{\Theta}^*$. The key tool for solving this problem is the characterization of AR equations representing MPUMs given in Theorem 6.5. We will call $\mathbf{e}_{n+1}$, defined by

$$\mathbf{e}_{n+1} := \Theta^*\left(\frac{d}{dt}\right)\mathbf{w}_{n+1}, \tag{8.6}$$

the *error* time series associated with $\mathbf{w}_{n+1}$. Given the function $f(x)$, we will denote by $f^{(j)}(\alpha)$ its jth derivative with respect to x, evaluated at $x = \alpha$.

Proposition 8.7: The error time series is purely exponential:

$$\mathbf{e}_{n+1} = \epsilon_{n+1}\exp_{\lambda_{n+1}}.$$

If $\mathbf{p}_{n+1}$ is given by (8.1), we have

$$\epsilon_{n+1} = \Theta^*(\lambda_{n+1})p_{n+1,0} \in \mathbf{k}^q. \tag{8.8}$$

If $\mathbf{p}_{n+1}$ is given by (8.2), we have

$$\epsilon_{n+1} = \Theta^*(\lambda_h)p_{h,\kappa_h} + \sum_{j=1}^{\kappa_h}\frac{\Theta^{*(j)}(\lambda_h)}{j!}p_{h,\kappa_h - j} \in \mathbf{k}^q. \tag{8.9}$$

Proof: For $\mathbf{a}, \mathbf{b} \in \mathbf{k}[s]$, the following identity holds:

$$\mathbf{a}\left(\frac{d}{dx}\right)\mathbf{b}\left(\frac{d}{dy}\right)\exp(xy) = \mathbf{b}\left(\frac{d}{dy}\right)\mathbf{a}\left(\frac{d}{dx}\right)\exp(xy).$$

The identity holds equally if $\mathbf{a}$ is a matrix polynomial and $\mathbf{b}$ a scalar polynomial. Consequently, for any matrix polynomial $\mathbf{a} = \theta$, the scalar polynomial $\mathbf{b} = s^\kappa$, $x = t$, and $y = \lambda$, the above formula yields

$$\theta\left(\frac{d}{dt}\right)t^\kappa\exp(\lambda t) = \frac{d^\kappa}{d\lambda^\kappa}[\theta(\lambda)\exp(\lambda t)].$$

This, in combination with (8.1), (8.2) implies (8.8) and (8.9). □

Let the first nonzero entry of ϵ_{n+1} from the top, be the rth. We can always normalize the error so that it has the form

$$\epsilon_{n+1} = \begin{pmatrix} 0 \\ 1 \\ \hat{\epsilon} \end{pmatrix} \in \mathbf{k}^q \qquad \text{where } \hat{\epsilon} \in \mathbf{k}^{q-r}, \tag{8.10}$$

[15] It might happen that $\hat{\mathbf{D}} = \mathbf{D}$.

and the 0 has dimension $r-1$. An AR equation representation of the error MPUM is

$$\Gamma_{n-1}(s) := \begin{pmatrix} I_{r-1} & 0 & 0 \\ 0 & s-\lambda_{n-1} & 0 \\ 0 & -\hat{\epsilon} & I_{q-r} \end{pmatrix} \in \mathbf{k}^{q\times q}[s]. \quad (8.11)$$

Clearly, $\Gamma_{n+1} = I$ whenever $\epsilon_{n-1} = 0$. Finally, let

$$\hat{\Theta}^* := \Gamma_{n+1}\Theta^* \in \mathbf{k}^{q\times q}[s]. \quad (8.12)$$

The following result solves the problem of recursively updating an AR representation of the MPUM.

Theorem 8.13: With the notation introduced above, $\hat{\Theta}^*$ is an AR equation representation of the MPUM $\hat{\Sigma}^*$ of $\hat{\mathbf{D}}$. It is obtained by updating the AR equation representation Θ^* of the MPUM Σ^* of $\mathbf{D}$ by means of premultiplication with Γ_{n+1}, as defined in (8.11)–(8.12).

Corollary 8.14: If Θ^* is row reduced with ordered row indexes $v_i \leq v_{i+1}$, where $v_i := deg\theta_i$, $i \in \underline{q}$, the updated $\hat{\Theta}^*$ defined by (8.12), is also row reduced.

Proof of 8.13: It is readily shown that $\hat{\Theta}^*$ is unfalsified. According to Theorem 6.5, therefore, in order to show that $\hat{\Theta}^*$ defined by (8.12) is an AR representation of the MPUM of $\hat{\mathbf{D}}$, it suffices to show that the nonunity invariant factors of $\hat{\Theta}^*$, defined by (8.12), are the same as those of $sI - \hat{F}$, where $\hat{F}$ is defined by (8.4), (8.5b). To prove this equality, we may assume without loss of generality, that

$$\lambda_i = 0, \qquad i \in \underline{n+1},$$

i.e., $\pi = s^N$, and consequently, all invariant factors are powers of s. We will also assume for simplicity, that $\kappa_i \leq \kappa_{i+1}$, $i \in \underline{n-1}$.

There are two cases to consider. If $\hat{F}$ is defined by (8.4), the (ordered) set of degrees of its nonunity invariant factors is:

$$\kappa_n, \kappa_{n-1}, \cdots, \kappa_1, 1. \quad (8.15)$$

If $\hat{F}$ is defined by (8.5b), on the other hand, the (possibly unordered) set of degrees of the nonunity invariant factors is:

$$\kappa_n, \cdots, \kappa_h + 1, \cdots, \kappa_1. \quad (8.16)$$

By (3.10a) Θ^* is uniquely determined up to left multiplication by a unimodular matrix. Recall the Smith decomposition of Section VII. We may thus assume without loss of generality that

$$\Theta^* = \Delta V, \Delta := \operatorname{diag}\left(\tilde{\Delta}, I_{q-n}\right)$$
$$\text{where } \tilde{\Delta} := \operatorname{diag}(s^{\kappa_n}, \cdots, s^{\kappa_1}), \quad (8.17)$$

while V is unimodular. Thus, the expression for (8.12) becomes

$$\hat{\Theta}^* := \Gamma_{n-1}\Delta V, \quad (8.18)$$

which implies that the invariant factors of $\hat{\Theta}^*$ are the same as the invariant factors of the product $\Gamma_{n+1}\Delta$, where Γ_{n+1} is defined by (8.11). Clearly, these are equal to the invariant factors of the product of two diagonal matrices, namely:

$$\operatorname{diag}(I_{r-1}, s, I_{q-r}) \cdot \operatorname{diag}\left(\tilde{\Delta}, I_{q-n}\right). \quad (8.19)$$

Combining (8.17) with (8.8), we conclude that, if (8.4) is satisfied, then $r > n$. Thus, from (8.19), the (ordered) set of degrees of the invariant factors of $\hat{\Theta}^*$ is the same as (8.15). On the other hand a simple calculation involving (8.17) and (8.9) shows that if (8.5) is satisfied, $r = i < n$. This implies that the (unordered) set of degrees of the invariant factors of $\hat{\Theta}^*$ is given by (8.16), as required.

The proof of the theorem is thus complete. □

For use in the next section, we now state without proof, the following result.

Auxiliary Proposition 8.20: Let $\Theta \in \mathbf{k}^{p\times q}[s]$, be row reduced with equal row degrees $\deg \theta_i = v$, $i \in \underline{p}$. Given is also $\epsilon \in \mathbf{k}^p$, $\epsilon \neq 0$. There exists a constant matrix $P \in \mathbf{k}^{p\times p}$, $\det P \neq 0$, such that

a) $\tilde{\Theta} := P\Theta$ remains row reduced with equal row indexes v

b) The given vector ϵ is transformed to $\tilde{\epsilon} := P\epsilon = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \end{pmatrix}$.

Corollary 8.21: Let $\epsilon = \epsilon_{n+1}$ defined by (8.10). We may assume without loss of generality that Θ^* is row reduced and the degree of the rth row of Θ^* is (strictly) less than the degree of the $(r+1)^{st}$ row:

$$\deg \theta_r < \deg \theta_{r+1}.$$

This corollary says that if the $(r+1)^{st}$ entry of the error ϵ_{n+1} is nonzero, and $\deg \theta_{r+1} = \deg \theta_r$, by preprocessing and reordering of the rows of Θ^*, the index r of the first nonzero entry of ϵ_{n-1} [defined by (8.10)] may be replaced by $r+1$. Eventually, repeating this procedure, the degree of the row of Θ^* corresponding to the first nonzero entry of the error ϵ_{n+1}, will be strictly greater than the degree of the next row. This can be achieved moreover, without affecting row reducedness.

Remarks 8.22:

a) Suppose that the new datum $\mathbf{w}_{n-1}$ is such that $\mathbf{p}_{n-1}$ has the form (1.2) with $\deg \mathbf{p} > 0$, but (8.2) is not satisfied, i.e., $(d\mathbf{p}_{n+1}/dt) \neq \mathbf{p}_i$, for any $i \in \underline{n}$, such that $\lambda_i = \lambda_{n+1}$. In this case the time series is new and the problem is equivalent to providing the following set of time series [cf. (4.1)]:

$$\mathbf{w}_{n+1,i-1} := \left[\frac{d}{dt} - \lambda_{n-1}\right]^{i-1} \mathbf{w}_{n+1},$$

where $i = 1, 2, \cdots, \deg \mathbf{p} + 1$.

b) Theorem 8.13 generalizes a result given in Willems [29, p. 289]. Therein, the error MPUM is expressed in a representation independent way. Here is a variation of that formula, valid for all λ_i:

$$\Gamma_{n+1}(s) := (s - \lambda_{n-1})I_q + I_q - \frac{\epsilon_{n+1}\epsilon_{n+1}^T}{\epsilon_{n+1}^T\epsilon_{n+1}}.$$

Formula 8.11 which we have adopted here, although not representation independent, has the advantage that it facilitates keeping track of the update of (1) the invariant factors, and (2) the row degrees of Θ^*, by virtue of Corollary 8.14. (1) was used above for the update of the AR equation representation of the MPUM of **D**, while (2) will be used in the next section for the update of the minimal-complexity controllable models of **D**.

c) From the proof of Theorem 8.13 we conclude that if $\lambda_{n+1} \neq \lambda_i, i \in \underline{n}$, or if $\lambda_{n+1} = \lambda_h$ for some $h \in \underline{n}$, and (8.1) is satisfied, *the number of invariant factors divisible by* $s - \lambda_{n+1}$ **increases** *by one*. If however, $\lambda_{n+1} = \lambda_h$ for some $h \in \underline{n}$, and (8.2) is satisfied, *the number of invariant factors divisible by* $s - \lambda_{n+1}$ **remains constant.** It is assumed of course, that the error MPUM is not zero (i.e., $p_{n+1} \neq 0$ and/or that the elements of $\hat{\mathbf{D}}$ satisfy the column-reducedness condition of Proposition 4.7-iii).

From (5.1), (5.4), (5.5) follows that the positive integer m_i defined in (4.6c), is equal to the number of Jordan blocks in F at the frequency λ_i, $i \in \underline{n}$. Thus, in the latter case, the number of nonunity invariant factors remains constant, while in the former, the number of these invariant factors will increase if

$$m = m_h,$$

for the subscript h defined above. As shown in the next section, the consequence of this fact is that the number of inputs of least-complexity controllable models might increase during an update.

d) Given the polynomial-exponential data sets $\mathbf{D}_1, \mathbf{D}_2$, let Θ_1^*, Θ_2^* be AR equation representations of the respective MPUMs, namely $\mathbf{B}_1^*, \mathbf{B}_2^*$. It follows from Proposition 3.10-e), that an AR equation representation $\hat{\Theta}^*$ of the MPUM $\hat{\mathbf{B}}^*$ of $\hat{\mathbf{D}} := \mathbf{D}_1 \cup \mathbf{D}_2$, is obtained as a *least common left multiple of* Θ_1^*, Θ_2^*:

$$\hat{\Theta}^* := lclm(\Theta_1^*, \Theta_2^*).$$

The *lclm* can be determined by computing a factorization of the rational function

$$\Theta_2^*(\Theta_1^*)^{-1} = A_1^{-1}A_2,$$

where the polynomial matrices A_1, A_2 are left co-prime. It follows that

$$\hat{\Theta}^* = A_2\Theta_1^* = A_1\Theta_2^*,$$

is the desired MPUM. Notice that by Proposition 3.10-b), whenever $\mathbf{B}_1^* \cap \mathbf{B}_2^* \neq 0$, Θ_1^* and Θ_2^* are *not* right coprime. In this section an explicit construction of the *lclm* is given for the case where $\mathbf{D}_1 = \mathbf{D}$ and $\mathbf{D}_2 = \mathbf{D}_{n+1}$. It turns out that $A_2 = \Gamma_{n+1}$. □

IX. Recursive Update of the Minimal-Complexity Controllable Models

Having described the update of AR equation representations of the MPUM Σ^*, our next goal is to describe the update of the *controllable* part of these AR equations. This will show how the family of *controllable minimal-complexity unfalsified models*, abbreviated C-MCUM and denoted [according to (3.23a)] by $\Sigma_{contr}^*(\alpha)$, is updated. Recall that α is the vector parameter parametrizing this family. The plan we are following is:

$$\hat{\mathbf{B}}^* \xrightarrow{\text{VIII}} \hat{\Theta}^* \xrightarrow{\text{IX}} \hat{\Theta}_{\text{contr}}^*(\alpha) \xrightarrow{\text{VII}} \hat{\Psi}_{\text{contr}}^*(\alpha).$$

The first arrow has been discussed in Section VIII; the second arrow will be discussed below; the last arrow will not be discussed explicitely; it can be completed using Formula 3.25 or Formula 7.2.

According to Section III-D-2) we may assume without loss of generality that

$$\Theta^* \text{ is row reduced with row indexes } v_1 \leq \cdots \leq v_q.$$

Assuming that condition iii) of Proposition 4.7 is satisfied, it readily follows that N, defined in (1.4), is equal to the sum of the v_i's:

$$N = v_1 + \cdots + v_r + \cdots v_q = n(\Sigma^*).$$

Recall 3.19: The following is a characterization of the MPUM and of the C-MCUM's of **D**.

Proposition 9.1: Let $\Sigma(\Theta)$, $\Theta \in \mathbf{k}^{g\times q}[s]$, be a model of **D**. Among all models of **D**:

a) Σ is the MPUM, denoted by Σ^*, if, and only if, $c(\Sigma)$ is minimal, i.e.,

$$g = q \text{ and } n(\Sigma) \text{ is minimal;}$$

b) Σ is a C-MCUM, denoted by Σ_{contr}^*, if, and only if, the invariant factors of Θ are unity, and $c(\Sigma)$ is minimal, i.e.,

$$g = p \text{ and } n(\Sigma) \text{ is minimal,}$$

where p is the number of unity invariant factors of Θ^*.

A consequence of the above proposition is that in order to update the MPUM we do not need to keep tract of the individual row degrees of Θ^*; only their sum N is needed. On the other hand, in order to update the C-MCUM's, we need to keep track of the individual degrees of the rows of Θ^*, due to the fact that $n(\Sigma_{contr}^*)$ is not readily computable from **D**. A consequence of Corollary 8.14 is the fact that the rth row degree of the updated AR equations is increased by one:

$$n(\hat{\Sigma}^*) = N + 1 = v_1 + \cdots + (v_r + 1) + \cdots + v_q.$$

Having obtained Θ^* and $\hat{\Theta}^*$, using Formula 7.1 we can compute all Θ_{contr}^* and $\hat{\Theta}_{contr}^*$. In order, however, to make the transition from Θ_{contr}^* to the updated $\hat{\Theta}_{contr}^*$ as explicit as possible, we need to keep track of the update of the *rows* and the *row degrees* of Θ^*. We proceed in two steps. In Section IX-A we will show how the family $\Sigma_{contr}^*(\alpha)$ can be parametrized, given Θ^* and *one* controllable model. The second step consists of simultaneously updating Θ^* and one controllable model (Section IX-B). Finally, Section IX-C is devoted to the proofs to the main results.

A. Affine Parametrization of Minimal-Complexity Controllable Models

First, some notation is introduced. Let p, as before, be the number of unity invariant factors of Θ^*; by Remark 6.10, we may assume without loss of generality that $p < q$. The index set

$$\mathbf{I} := \{i_j \in \underline{q},\ j \in \underline{p}\}, \tag{9.2}$$

and the associated partial permutation matrix:

$$\Pi := \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_p \end{pmatrix} \in \mathbf{k}^{p \times q} \qquad \text{where } \pi_j := \xi_{i_j}^{16}, \tag{9.3}$$

are defined. A key result is the following. Its proof is given in Section IX-C.

Main Lemma 9.4: An AR equation representation Θ^* of the MPUM Σ^* of $\mathbf{D}$, which is computed in Section VI and updated in Section VIII, can always be chosen to satisfy the following properties simultaneously.

a) It is row reduced with row indexes ordered in increasing order.

b) There exists an index set $\mathbf{I}$, such that $\Pi\Theta^*$, has full row rank p for all $\lambda \in \mathbf{C}$, where p is the number of unity invariant factors of Θ^*, and the sum $v_{i_1} + \cdots + v_{i_p}$ *is minimal.*

The above result implies that *one* controllable model of minimal complexity is composed of the rows of an appropriate Θ^*, which are indexed by the set I:

$$\Theta^*_{\text{contr}} = \Pi\Theta^*.$$

Given this model, all other minimal-complexity controllable models can be obtained by adding to each row with index $i \in \mathbf{I}$, polynomial multiples of all rows with index j satisfying: $j \in \underline{q}$, $j \notin \mathbf{I}$, and $\deg v_j \le \deg v_i$.

Our goal in the sequel is to determine this parametrization explicitly. To simplify the presentation, we will assume for the remaining of Section IX-A, without loss of generality, that the rows of Θ^* have been permuted so that

$$\mathbf{I} = \underline{p},\quad v_i \le v_{i-1},\quad i \in \underline{p-1},$$
$$v_{p-j} \le v_{p-j+1},\quad j \in \underline{m-1}; \tag{9.5}$$

as before, $p + m = q$. Θ^*_{contr} is thus composed of the first p rows of Θ^*:

$$\Theta^*_{\text{contr}} = (I_p \quad 0)\Theta^*.$$

The degrees of these p rows are ordered; so are the degrees of the m subsequent rows. The family of C-MCUM's defined by (3.23a) can now be parametrized as follows: $\Sigma^*_{contr}(\alpha) = \Sigma(\Theta^*_{contr}(\alpha))$, where

$$\Theta^*_{\text{contr}}(\alpha) := \big(I_p \quad A(\alpha)\big)\Theta^* \tag{9.6a}$$

A is a matrix, whose entries are arbitrary polynomials of appropriate degree:

$$A(\alpha) \in \mathbf{k}^{p \times m}[s],$$
$$\deg a_{ij} = v_i - v_{p-j} \qquad \text{if } v_i \ge v_{p+j} \text{ and } a_{ij} = 0 \text{ otherwise.} \tag{9.6b}$$

The parameter α consists of the coefficients of each one of the nonzero polynomials a_{ij}, it has κ entries:

$$\alpha \in \mathbf{k}^{\kappa}, \qquad \kappa := \Sigma(v_i - v_{p+j} + 1),$$

where the sum is taken over all indexes $i \in \underline{p}$, $j \in \underline{m}$, such that $v_i \ge v_{p+j}$. By construction, $\Theta^*_{constr}(\alpha)$ is row reduced with the same row degrees as Θ^*_{contr}. Moreover, since $A(\alpha)$ is linear in the parameter α, the Parametrization 9.6 is *affine*. For controllability, according to Section III-E certain values of α in $\mathbf{k}^{\kappa}$ have to be excluded. The condition which has to be satisfied is

$$\operatorname{rank} \Theta^*_{\text{contr}}(\lambda, \alpha) = p, \qquad \lambda \in \mathbf{C}.$$

Since however $\det \Theta^*(\lambda) \ne 0$, for $\lambda \ne \lambda_i$, $i \in \underline{n}$, this condition is equivalent to

$$\operatorname{rank} \Theta^*_{\text{contr}}(\lambda_i, \alpha) = \operatorname{rank}\big(I_p \quad A(\lambda_i, \alpha)\big)\Theta^*(\lambda_i) = p, \qquad i \in \underline{n}. \tag{9.7}$$

Because of (9.4b), there exist constant matrices $M_i \in \mathbf{k}^{q \times p}$, $N_i \in \mathbf{k}^{m \times p}$, such that

$$\Theta^*(\lambda_i)M_i = \begin{pmatrix} I_p \\ N_i \end{pmatrix}, \qquad i \in \underline{n}.$$

The following n multi-linear functions of α will be needed below:

$$f_i(\alpha) := \det\big(I_p + A(\lambda_i, \alpha)N_i\big) = \det(I_m + N_i A(\lambda_i, \alpha)), \qquad i \in \underline{n}. \tag{9.8a}$$

If μ is the maximum between p and the number of nonzero columns of A, the f_i's are actually μ-linear functions of the parameter α. A simply argument shows that (9.7) is equivalent to the condition

$$f_i(\alpha) \ne 0, \qquad i \in \underline{n}. \tag{9.8b}$$

The following special cases are worth noting. Recall the Assumption 9.5. i) If $v_p < v_{p-1}$, $A(\alpha) = 0$; thus there exists a unique C-MCUM. ii) If $v_{p-1} \le v_p < v_{p-2}$, the matrix A has one nonzero column, and hence constraints (9.8) are *linear*. The same holds true if $p = 1$. iii) If $v_{p+1} \le v_{p+2} \le v_p < v_{p-3}$, the parameter matrix A has two nonzero columns and thus constraints (9.8) are in general *bilinear*. The same holds true if $p = 2$. There follows an example illustrating some of the above considerations. More instances of computation of the family $\Sigma^*_{contr}(\alpha)$ can be found in Section XI.

[16] We use ξ_l to denote the lth row unit q-vector.

Example: Let **D** be such that $\lambda_1 = 0$, $\lambda_2 = -1$, and the behavior of the MPUM is $\mathbf{B}(\Theta^*)$, where

$$\Theta^* = \begin{pmatrix} sI_2 & I_2 \\ 0 & (s+1)I_2 \end{pmatrix} \in \mathbf{k}^{4\times 4}[s].$$

Assumptions (9.5) are fulfilled. Following (9.6), the family of all minimal-complexity controllable models $\Sigma(\Theta^*_{contr}(\alpha))$ or **D** can be parametrized as follows:

$$\Theta^*_{\mathrm{contr}}(s,\alpha) = [sI_2 \quad (s+1)A(\alpha) + I_2] \in \mathbf{k}^{2\times 4}[s],$$

$$A(\alpha) = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{pmatrix} \in \mathbf{k}^{2\times 2}.$$

It follows that $N_1 = I_2$ and $N_2 = 0$. Hence, $f_1 = \det(I_2 + A(\alpha))$ and $f_2 = 1$. Therefore by (9.8), the parameter vector $\alpha := (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ has to satisfy the bilinear constraint $(\alpha_1 + 1)(\alpha_4 + 1) - \alpha_2\alpha_3 \neq 0$. □

The above considerations are summarized in the following theorem.

Theorem 9.9: Affine parametrization of minimal-complexity controllable models of **D**. With (9.5) holding, given one minimal-complexity controllable model $\Sigma^*_{contr} = \Sigma(\Theta^*_{contr})$, all others $\Sigma^*_{contr}(\alpha) = \Sigma(\Theta^*_{contr}(\alpha))$, are obtained by Formula 9.6. The coefficients of each nonzero entry of the polynomial matrix A, make up the parameter vector α; the range of α is $\mathbf{k}^{\kappa}$, with the exception of the μ-linear surfaces defined by (9.8).

We thus conclude that in order to update the family $\Sigma^*_{contr}(\alpha)$, we first need to update one of its elements, say, Σ^*_{contr} to $\hat{\Sigma}^*_{contr}$, and then apply the above theorem. The next section is thus devoted to the computation of the update $\hat{\Sigma}^*_{contr}$ of Σ^*_{contr}.

B. Recursive Update of a C-MCUM

Recall (8.10)–(8.13). Corollary 8.14 asserts that, with Γ_{n+1} defined by (8.11), the row reducedness of Θ^* implies the row reducedness of $\hat{\Theta}^*$. Furthermore, by Corollary 8.21, we may assume without loss of generality that $\deg \theta_r < \deg \theta_{r-1}$.[17] Let

$$\Delta := \begin{pmatrix} I_{r-1} & & \\ & 1 & \\ & \alpha & I_{q-r} \end{pmatrix} \in \mathbf{k}^{q\times q},\ \alpha \in \mathbf{k}^{q-r}. \quad (9.10a)$$

In the sequel instead of (8.12), the following modified form of the updated MPUM will be used

$$\hat{\Theta}^*(s) := \Delta\Gamma_{n-1}(s)\Theta^*(s). \quad (9.10b)$$

Due to 8.21 $\hat{\Theta}^*$ just defined is also row reduced. Given a matrix M^* with q rows and a set $\mathbf{I} \subset \underline{q}$, $M_{\mathbf{I}}$ denotes the submatrix of M^* composed of those rows which are indexed by **I**.

Here is the *key result* of this section. It says that the update can be arranged in such a way that the rows indexed by **I** have full row rank for all frequencies, with the possible exception of the $(n+1)^{st}$, in which case the rank might drop by one. Its proof is given in Section IX-C below.

[17] Recall from (8.10) that $r := \min\{i \in \underline{q}\colon \epsilon_i \neq 0\}$.

Proposition 9.11: With the notation introduced above and **I** as in (9.2), $\hat{\Theta}^*$ defined by (9.10) satisfies for all α

$$\operatorname{rank} \hat{\Theta}_{\mathbf{I}}(\lambda_{n+1}) \geq p - 1 \quad (9.11a)$$

while there exists $\alpha \in \mathbf{k}^{q-r}$, such that

$$\operatorname{rank} \hat{\Theta}_{\mathbf{I}}(\lambda) = p, \quad \forall \lambda \neq \lambda_{n+1}. \quad (9.11b)$$

We are now ready to define the indexes k, l which play a central role in the C-MCUM update.

$$l := \min\Big\{ i \in \mathbf{I} \subset \underline{q}\colon \hat{\theta}_i(\lambda_{n+1}) \text{ linearly dependent on } \hat{\theta}_j(\lambda_{n+1}),\, j < i,\, j \in \mathbf{I} \Big\} \quad (9.11c)$$

$$k := \min\left\{ i \notin \mathbf{I}, i \in \underline{q}\colon \operatorname{rank}\begin{bmatrix} \hat{\Theta}_{\mathbf{I}-\{l\}}(\lambda_{n-1}) \\ \hat{\theta}_i(\lambda_{n-1}) \end{bmatrix} = p \right\}. \quad (9.11d)$$

Both of the above indexes might fail to exist. From the proposition it follows that the index l does not exist, iff $r \notin \mathbf{I}$ and $\theta_r(\lambda_{n+1})$ is linearly independent from the rows of $\Theta_{\mathbf{I}}$; moreover, if $r \in \mathbf{I}$, $l = r$. If l exists, k does *not* exist, iff the number of non-unity invariant factors of the MPUM increases during the update; as already mentioned, this will happen if the new measurement is linearly independent and there is a majority of measurements in $\hat{\mathbf{D}}$ at the frequency λ_{n+1} (see also Remark 9.20b). From Proposition 9.11 and the definition (9.11c, d) of the indexes l, k, follows the classification of the C-MCUM update problem. For the proof, see Section IX-C.

Lemma 9.12: The update of a C-MCUM. Let $\Theta^*_{\mathrm{contr}}$ be composed of those rows of Θ^* indexed by the elements of **I** defined in (9.2). With the updated MPUM $\hat{\Theta}^*$ given by (9.10), a C-MCUM $\hat{\Theta}^*_{\mathrm{contr}}$ is given as follows:

a) If l does not exist, $\hat{\mathbf{I}} = \mathbf{I}$ and $\hat{\Theta}^*_{\mathrm{contr}} = \hat{\Theta}_{\mathbf{I}}$.

b) If l exists but k does not exist, $\mathbf{I} = \mathbf{I} - \{l\}$ and $\hat{\Theta}^*_{\mathrm{contr}} = \hat{\Theta}_{\mathbf{I}-\{l\}}$.

c) If both l and k exist, $\hat{\mathbf{I}} = \mathbf{I} - \{l\} \cup \max\{l, k\}$ and $\hat{\Theta}^*_{\mathrm{contr}} = [\hat{\Theta}_{\mathbf{I}-\{l\}} \quad a\hat{\theta}_l + \hat{\theta}_k]$, where the constant $a \in \mathbf{k}$ is chosen so that $\hat{\Theta}^*_{\mathrm{contr}}$ has full row rank p for all frequencies λ.

Corollary 9.14: Let the set of degrees δ of the C-MCUM Σ^*_{contr} be:

$$\delta := \{v_{i_1}, \cdots, v_{i_p}\}.$$

The set of degrees $\hat{\delta}$ of the updated C-MCUM Σ^*_{contr} is:

a) Either, $\hat{\delta} = \delta$;

b) Or, $\hat{\delta} = \delta_l := \delta - \{v_l\}$.

c) Or, $\hat{\delta} = \delta_l \cup \max\{v_l + 1, v_k\}$.

Remark 9.15: Although (9.12) provides the solution of the C-MCUM update problem, this is *not* the end of the story. The reason is that in order to be ready for the next update, we need to account for the remaining rows (i.e., those rows which are not part of the C-MCUM considered). The need for keeping track of these rows is readily seen by considering Definition 9.11; the kth row which enters in the update of the C-MCUM is *not* part of the C-MCUM at the previous stage. □

To achieve this goal, the definition of a few more quantities is needed. Recall Definition 9.3 of Π.

$$\Pi_i \in \mathbf{k}^{(p-1)\times q}\text{: contains the same rows as } \Pi \text{ with } \xi_i \text{ eliminated}, \quad (9.16a)$$

$$\Pi_{ij} \in \mathbf{k}^{p\times q}\text{: contains the same rows as } \Pi \text{ with } \xi_i \text{ replaced with } \xi_j. \quad (9.16b)$$

Finally, let $A \in \mathbf{k}^{q\times q}$, $\det A \neq 0$, be defined as follows:

$$\text{if } k \text{ does not exist: } A := I_q; \quad (9.17a)$$

if l and k exist and $l > k$ then $(A)_{i,i} := 1$, $i \in \underline{q}$, $i \neq l$ and

$$(A)_{l,k} := 1, (A)_{l,l} := a; \quad (9.17b)$$

if l and k exist and $l < k$ then $(A)_{i,j} := 1$, $i \in \underline{q}$, $i \neq l$ and

$$(A)_{k,l} := a, (A)_{k,k} := 1 \quad (9.17c)$$

all other entries in (9.17b, c) are zero. We are now ready to state the main result. Cases a), b), c) in Lemma (9.12) correspond to the three cases listed in the theorem below. Recall (8.10)–(8.12).

Theorem 9.18: Simultaneous update of the MPUM and of a C-MCUM. With the notation introduced above, an AR equation representation of the C-MCUM $\hat{\Sigma}^*_{contr}$, and of the MPUM $\hat{\Sigma}^*$, of $\hat{\mathbf{D}}$, is

$$\hat{\Theta}^*_{\text{contr}} = \hat{\Pi}\hat{\Theta}^*, \quad \hat{\Theta}^* := A\Gamma_{n+1}\Theta^*, \quad (9.19)$$

where $\hat{\Pi}$ is defined as follows:

a) If l does not exist $\hat{\Pi} := \Pi$.

b) If l exists but k does not exist, $\hat{\Pi} := \Pi_l$.

c) If l and k exist, $\hat{\Pi} := \Pi$ if $l > k$, and $\hat{\Pi} := \Pi_{lk}$ for $r < k$.

Remarks 9.20:

a) In the above considerations we concentrated our attention the class of minimal-complexity controllable models $\Sigma^*_{contr}(\alpha)$, introduced in (3.23a). It is now a straightforward matter to look for the I/O controllable models of minimal complexity, denoted by $\Sigma^*_{I/O}(\beta)$, and for the *smooth* I/O controllable models of minimal complexity, denoted by $\Sigma^*_{pr}(\gamma)$, of the data $\mathbf{D}$ [see (3.23b, c)]. Property b) of main Lemma 9.4 has to be appended by requiring, in addition, that

$$\det \mathbf{T} \neq 0 \quad \text{and} \quad \deg\det \mathbf{T} = v_{i_1} + \cdots + v_{i_p},$$

respectively. That such properties can always be satisfied, can be proved recursively by appropriately modifying the proof of (9.4) given in Section IX-C below.

b) The number of unity invariant factors of Θ^*, which we have denoted by p above, is actually equal to the number of output variables of the I/O system associated with any C-MCUM as discussed in remark a) above. Consequently, $m := q - p$, is the corresponding number of input variables.

It is interesting to note in this context, that according to the above results, the number of inputs of C-MCUMs will either remain constant, or will *increase* during the updating procedure. This happens in case b) of Theorem 9.18, which in turn corresponds to case b) of Corollary 9.14. In other words, whenever we want to model an additional polynomial-exponential time series having frequency λ_{n-1} equal to the frequency of a time series already modeled, but which has different direction in $\mathbf{k}^q$ [that is (8.1) is fulfilled)], if there is a majority of time series at the frequency λ_{n+1} in $\hat{\mathbf{D}}$, *then* the number of input variables of the model increases at the expense of the number of output variables [see also Remark 8.19-c)].

c) A nonrecursive result similar to Lemma 9.4 was proved in Antoulas, Ball, Kang, and Willems [8, lemma 3.10], for the special case of the interpolation problem. Moreover, Theorem 9.18 above, corresponds to Theorem 4.13 of the same reference.

d) In the case where $q = 2$, and $m = p = 1$, Corollary 9.14 is simplified as follows. Let the complexity of $\Sigma^*_{contr}, \hat{\Sigma}^*_{contr}$ be $v_1, \hat{v}_1$, respectively. It follows that $\hat{v}_1$ can take one of the following values:

$$v_1; \text{ or } v_2 \text{ provided } v_2 > v_1; \text{ or } v_1 + 1.$$

Notice that the value $v_2 + 1$ is *not* allowed.

e) Corollary 9.14 provides the generalization of the corresponding result for the recursive update in the realization problem given in Antoulas [1, proposition 6.18]. In Antoulas and Anderson [2, proposition 3.6] the possible degrees of a recursive update in the scalar interpolation problem are derived; this latter result is formally *identical* to the result quoted in remark d) above. Notice also that in the special case of the recursive realization update, in addition to the value $v_2 + 1$, the value $v_1 + 1$, is *not* allowed either. □

C. Proof of Proposition 9.11 and Lemma 9.12

As already mentioned, the key result for the recursive update of a C-MCUM is Proposition 9.11 whose proof is given below. The first consequence is Lemma 9.12 which classifies the update problem into three cases; its proof is also given below. The proofs of Theorem 9.18 and Lemma 9.4 are immediate corollaries of (9.11) and (9.12) and are omitted.

Proof of Proposition 9.11: We will prove the proposition assuming that $r = 1$ and that the new measurement satisfies condition (8.1). The proof for $r > 1$, and/or in case condition (8.2) is satisfied, follows along the same lines. Let θ_1 denote the first row of Θ^* and $\Theta^* := (\theta_1 \quad \Theta_2)$. From (9.10) the updated MPUM is

$$\hat{\Theta}^*(s) = \begin{pmatrix} (s - \lambda_{n-1})\theta_1 \\ \Theta_2 + [(s - \lambda_{n-1})\alpha - \hat{e}]\theta_1 \end{pmatrix}.$$

Case 1: $r \in \mathbf{I}$ (i.e., the first row of Θ^* belongs to the C-MCUM at the n^{th} step). For appropriate $x(\alpha) \in \mathbf{k}^{p-1}$

$$\hat{\Theta}_{\mathbf{I}}(s) = \begin{pmatrix} s - \lambda_{n-1} & 0 \\ x(\alpha) & I_{p-1} \end{pmatrix}\Theta_{\mathbf{I}}(s).$$

It readily follows that conditions (9.11a, b) are satisfied, the former with equality sign; this holds for all α.

Case 2: $r \notin \mathbf{I}$. The following expressions holds [recall (8.8), (8.10)]:

$$\hat{\Theta}_{\mathrm{I}}(s) = \Theta_{\mathrm{I}}(s) + [(s - \lambda_{n-1})\alpha_{\mathrm{I}} - \bar{\epsilon}_{\mathrm{I}}]\theta_l(s) \quad (9.21a)$$

$$\begin{pmatrix} 1 \\ \hat{\epsilon} \end{pmatrix} = \Theta^*(\lambda_{n+1})p_{n+1,0} = \begin{pmatrix} \theta_l(\lambda_{n+1}) \\ \Theta_2(\lambda_{n+1}) \end{pmatrix} p_{n+1,0} \quad (9.21b)$$

which implies

$$\hat{\epsilon}_{\mathrm{I}} = \Theta_{\mathrm{I}}(\lambda_{n-1})p_{n-1,0}. \quad (9.21c)$$

Evaluating (9.21a) at $s = \lambda_{n+1}$ and using (9.21b) we obtain

$$\hat{\Theta}_{\mathrm{I}}(\lambda_{n+1}) = \Theta_{\mathrm{I}}(\lambda_{n+1})\left[I_q - p_{n+1,0}\theta_l(\lambda_{n+1})\right].$$

The expression in brackets has rank $n - 1$ and its left kernel is spanned by $\theta_l(\lambda_{n-1})$. Hence, $\hat{\Theta}_{\mathrm{I}}(\lambda_{n+1})$ has full row rank p iff $\theta_l(\lambda_{n+1})$ is linearly independent from the rows of $\Theta_{\mathrm{I}}(\lambda_{n+1})$. Otherwise, the rank is $p - 1$. This shows the validity of (9.11a) in case $r \notin \mathbf{I}$.

To show (9.11b), we first notice that if $\theta_l(\lambda)$, $\lambda \neq \lambda_{n+1}$, is linearly independent from the rows of $\Theta_{\mathrm{I}}(\lambda)$, then $\hat{\Theta}_{\mathrm{I}}(\lambda)$ has full row rank p for all choices of α. There remains to show (9.11b) in case the relationship

$$\theta_l(\lambda) = \gamma_\lambda \Theta_{\mathrm{I}}(\lambda),\ \gamma_\lambda \in \mathbf{k}^{1\times p}$$

holds, for some frequency $\lambda \neq \lambda_{n-1}$. Evaluating (9.21a) at $s = \lambda$ and substituting the above expression we obtain

$$\hat{\Theta}_{\mathrm{I}}(\lambda) = \left[I_p + (\lambda - \lambda_{n-1})\alpha_{\mathrm{I}}\gamma_\lambda - \epsilon_{\mathrm{I}}\gamma_\lambda\right]\Theta_{\mathrm{I}}(\lambda).$$

By construction, $\Theta_{\mathrm{I}}(\lambda)$ has full row rank p. Hence, $\hat{\Theta}_{\mathrm{I}}(\lambda)$ has full row rank p iff the matrix in brackets has nonzero determinant. Using the fact that $\det[I + AB] = \det[I + BA]$ we obtain the following constraints on the elements of the vector α:

$$\gamma\lambda[(\lambda - \lambda_{n+1})\alpha_{\mathrm{I}} - \epsilon_{\mathrm{I}}] + 1 \neq 0. \quad (9.11e)$$

This shows that for (9.11b) to hold, the components of α indexed by $\mathbf{I}$ have to avoid the hyperplanes (9.11e). This completes the Proof of 9.11. ■

Proof of Lemma 9.12: Cases a) and b) are immediate consequences of Proposition 9.11 and definition (9.11c, d) of the indexes l and k. There remains to show that in case c), the constant a can be chosen so that $\hat{\Theta}^*_{\mathrm{contr}}$ has full row rank p. By construction, for $s = \lambda_{n+1}$, $\hat{\theta}(\lambda_{n+1})$ is linearly dependent on $\hat{\Theta}_{\mathrm{I}-\{l\}}(\lambda_{n-1})$. Hence $\hat{\Theta}^*_{\mathrm{contr}}(\lambda_{n+1})$ has full row rank irrespective of the value of a. Let $s = \lambda$, $\det \hat{\Theta}^*(\lambda) = 0$, be such that

$$\hat{\theta}_k(\lambda) = \alpha_\lambda \hat{\Theta}_{\mathrm{I}-\{l\}}(\lambda) + \beta_\lambda \hat{\theta}_l(\lambda),\ \alpha_\lambda \in \mathbf{K}^{1\times p-1},\ \beta_\lambda \in \mathbf{k}.$$

It readily follows that the constant a has to be such that $a + \beta_\lambda \neq 0$. This completes the Proof of 9.12. ■

X. System-Theoretic Interpretations and Linear Fractions

Consider the system $\Sigma := (\mathbf{R}, \mathbf{k}^{2q}, \mathbf{B})$ with manifest variables

$$\begin{pmatrix} \bar{\mathbf{w}} \\ \mathbf{w} \end{pmatrix} \in \mathbf{B} \qquad \text{where } \mathbf{w} := \begin{pmatrix} \mathbf{u} \\ \mathbf{y} \end{pmatrix},\quad \bar{\mathbf{w}} := \begin{pmatrix} \bar{\mathbf{y}} \\ \bar{\mathbf{u}} \end{pmatrix} \in (\mathbf{k}^q)^{\mathbf{R}} \quad \text{and}$$

$$\mathbf{u}, \bar{\mathbf{y}} \in (\mathbf{k}^m)^{\mathbf{R}},\quad \mathbf{y}, \bar{\mathbf{u}} \in (\mathbf{k}^p)^{\mathbf{R}}. \quad (10.1)$$

As before $q = m + p$. The behavior $\mathbf{B}$ is

$$\mathbf{B} := \ker\left(I_q \quad -\Theta\left(\frac{d}{dt}\right)\right)$$

$$\text{where } \Theta^* := \begin{pmatrix} -\mathbf{R} & \mathbf{U} \\ \mathbf{Q} & -\mathbf{T} \end{pmatrix} \in \mathbf{k}^{q\times q}[s],$$

$$\mathbf{T} \in \mathbf{k}^{p\times p}[s],\quad \det\mathbf{T} \neq 0. \quad (10.2)$$

Identifying $\mathbf{u}, \bar{\mathbf{u}}$ as inputs and $\mathbf{y}, \bar{\mathbf{y}}$ as outputs, because of the nonsingularity of $\mathbf{T}$, by (3.13b), (3.14) Σ is an I/O system. It is sometimes referred to as a (linear) *two-port*; Θ is known as the *chain parameter* matrix relating $\bar{\mathbf{w}}$ and $\mathbf{w}$:

$$\bar{\mathbf{w}} = \Theta\left(\frac{d}{dt}\right)\mathbf{w}.$$

This two-port is depicted in Fig. 1. Following the results of Section III-E, since the rank of $(I_q \quad -\Theta(\lambda))$ is equal to q for all complex frequencies λ, Σ is controllable. Hence, it can be equivalently described in terms of the transfer function Z which relates the inputs $\begin{pmatrix} \bar{\mathbf{u}} \\ \mathbf{u} \end{pmatrix}$ to the outputs $\begin{pmatrix} \bar{\mathbf{y}} \\ \mathbf{y} \end{pmatrix}$:

$$Z = -\begin{bmatrix} \mathbf{U}\mathbf{T}^{-1} & \mathbf{R} - \mathbf{U}\mathbf{T}^{-1}\mathbf{Q} \\ \mathbf{T}^{-1} & -\mathbf{T}^{-1}\mathbf{Q} \end{bmatrix}.$$

This is known in the literature (see e.g., Antoulas [4]) as the *transfer parameter* matrix of the two-port associated with the above chain parameters.

We now define two more systems: $\bar{\Sigma} := (\mathbf{R}, \mathbf{k}^q, \bar{\mathbf{B}})$, with behavior

$$\bar{\mathbf{B}} := \ker\Gamma$$

$$\text{where } \Gamma := (\bar{\mathbf{Q}} \quad -\bar{\mathbf{T}}) \in \mathbf{k}^{p\times q}[s],\quad \bar{\mathbf{T}} \in \mathbf{k}^{p\times p}[s], \quad (10.3a)$$

and $\hat{\Sigma} := (\mathbf{R}, \mathbf{k}^q, \hat{\mathbf{B}})$, with behavior

$$\hat{\mathbf{B}} := \ker\Gamma\Theta. \quad (10.3b)$$

If we interconnect Σ and $\bar{\Sigma}$ by imposing the constraint

$$\bar{\mathbf{w}} \in \bar{\mathbf{B}}, \qquad \text{that is } \Gamma\left(\frac{d}{dt}\right)\bar{\mathbf{w}} = 0,$$

it follows that

$$\mathbf{w} \in \hat{\mathbf{B}}, \qquad \text{that is } (\Gamma\Theta)\left(\frac{d}{dt}\right)\mathbf{w} = 0.$$

The converse holds also, i.e., if $\mathbf{w} \in \hat{\mathbf{B}}$ then $\bar{\mathbf{w}} := \Theta\mathbf{w} \in \bar{\mathbf{B}}$. Such an interconnection is called a *cascade* interconnection of the two-port system Σ with the (one-port) system $\bar{\Sigma}$; it is shown in Fig. 2. The overall system $\hat{\Sigma}$ has input $\mathbf{u}$ and output $\mathbf{y}$. From the above expressions we get

$$\Gamma\Theta = (-(\bar{\mathbf{Q}}\mathbf{R} + \bar{\mathbf{T}}\mathbf{Q}) \quad (\bar{\mathbf{Q}}\mathbf{U} + \bar{\mathbf{T}}\mathbf{T})) \in \mathbf{k}^{p\times q}[s].$$

If $\Gamma\Theta$ evaluated at every complex frequency has constant rank q, and $\det(\bar{\mathbf{Q}}\mathbf{U} + \bar{\mathbf{T}}\mathbf{T}) \neq 0$, $\hat{\Sigma}$ is a controllable I/O system. It can be equivalently described by means of the transfer function

$$\hat{Z} = (\bar{\mathbf{Q}}\mathbf{U} + \bar{\mathbf{T}}\mathbf{T})^{-1}(\bar{\mathbf{Q}}\mathbf{R} + \bar{\mathbf{T}}\mathbf{Q}). \tag{10.4a}$$

If, in addition, $\det\bar{\mathbf{T}} \neq 0$, $\bar{\Sigma}$ is also a controllable I/O system, and $\hat{Z}$ can be written as

$$\hat{Z} = (\bar{Z}\mathbf{U} + \mathbf{T})^{-1}(\bar{Z}\mathbf{R} + \mathbf{Q}) \qquad \text{where } \bar{Z} := \bar{\mathbf{T}}^{-1}\bar{\mathbf{Q}}. \tag{10.4b}$$

In this case we say that $\hat{Z}$ is expressed as a linear (left) fraction involving Θ^* and $\bar{Z}$.

The cascade interconnection of n two-ports is readily defined. Let $\Sigma_i := (\mathbf{R}, \mathbf{k}^{2q}, \mathbf{B}_i)$, with $\mathbf{B}_i := \ker(I \quad -\Theta_i)$, and manifest variables $\begin{pmatrix}\bar{\mathbf{w}}_i\\ \mathbf{w}_i\end{pmatrix}$:

$$\bar{\mathbf{w}}_i = \Theta_i\left(\frac{d}{dt}\right)\mathbf{w}_i, \qquad i \in \underline{n}.$$

If the interconnection constraint $\bar{\mathbf{w}}_i = \mathbf{w}_{i+1}$, $i \in n-1$, is imposed, the overall system Σ is a two-port relating $\bar{\mathbf{w}} := \bar{\mathbf{w}}_n$ and $\mathbf{w} := \mathbf{w}_1$:

$$\bar{\mathbf{w}} = \Theta\left(\frac{d}{dt}\right)\mathbf{w} \qquad \text{where } \Theta := \Theta_n\Theta_{n-1}\cdots\Theta_2\Theta_1.$$

Σ is the *cascade* interconnection of the two-ports Σ_i, as depicted is Fig. 3.

The above considerations readily apply to our polynomial-exponential modeling problem. Recall the MPUM $\mathbf{B}^*$ of the data $\mathbf{D}$, and the AR equation representation Θ^* derived in Section VI. For any choice of the input and output variables, the rows of Θ^* can be permuted so that the nonsingularity of $\mathbf{T}$ in the partitioning (10.2) is assured. Then according to the above remarks, Θ^* can be interpreted as a two-port. Moreover the parametrization of all models given by (3.11), (3.12) can be interpreted as a cascade interconnection of this two-port denoted by Σ, with the terminating system denoted by $\hat{\Sigma}$, as shown in Fig. 2. Since Θ^* depends only on the data while Γ is arbitrary, we sometime refer to Θ^* as a *generating system* for *all* models of $\mathbf{D}$.

Following the discussion of Section XIII, the MPUM of $\mathbf{D}$ can be recursively constructed from the successive error MPUMs. Let $\hat{\Gamma}_i$ by AR equations for these error MPUMs. An AR equation Θ^* for the overall MPUM is given by their product:

$$\Theta^* = \cdots\hat{\Gamma}_n\cdots\hat{\Gamma}_2\hat{\Gamma}_1.$$

The recursive solution of the polynomial-exponential time series modeling problem, has therefore the interpretation of the cascade interconnection Σ of the two-ports Σ_i, each defined in terms of the corresponding $\hat{\Gamma}_i$, as depicted in Fig. 3.

The parametrization of *all* controllable I/O models (including minimal-complexity ones) can be expressed in terms of linear fractions as in (10.4a, b). The same holds true for the *recursive update* of the controllable I/O models. It should be mentioned that the parametrization of minimal-complexity controllable solutions, as well as their recursive update by means of linear fractions, was first derived for the realization problem by Antoulas [1, Theorems 3.5 and 5.8]. For an overview, see also Antoulas [4].

XI. Examples

Example A: Consider the data set $\mathbf{D} := \{\mathbf{w}_1 = \mathbf{p}_1 \in (\mathbf{R}^4)^{\mathbf{R}}, \mathbf{w}_2 = \mathbf{p}_2 \in (\mathbf{R}^4)^{\mathbf{R}}\}$, with $\lambda_1 = \lambda_2 = 0$, $\mathbf{p}_1(t) := p_{1,0}t + p_{1,1}$ i.e. $\kappa_1 = 2$, and $\mathbf{p}_2(t) := p_{2,0}(t^3/3!) + p_{2,1}(t^2/2!) + p_{2,2}(t/1!) + p_{2,3}$ i.e. $\kappa = 4$;

$$p_{1,0} := \begin{pmatrix}0\\0\\0\\4\end{pmatrix}, \quad p_{1,1} := \begin{pmatrix}0\\0\\0\\-3\end{pmatrix}, \quad p_{2,0} := \begin{pmatrix}2\\0\\0\\0\end{pmatrix},$$
$$p_{2,1} := \begin{pmatrix}3\\-2\\0\\0\end{pmatrix}, \quad p_{2,2} := \begin{pmatrix}0\\-3\\2\\0\end{pmatrix}, \quad p_{2,3} := \begin{pmatrix}0\\0\\3\\0\end{pmatrix}.$$

With $N := \kappa_1 + \kappa_2 = 6$, the MPUM of $\mathbf{D}$ is $\Sigma(\Theta^*)$ where

$$\Theta^*(s) = \begin{pmatrix}s & 1 & 0 & 0\\ 0 & s & 1 & 0\\ 0 & 0 & s^2 & 0\\ 0 & 0 & 0 & s^2\end{pmatrix} \Rightarrow c^* = (0, 6).$$

In this case, there is a unique minimal-complexity controllable model $\Sigma(\Theta^*_{contr})$ [cf. (3.23a)], where

$$\Theta^*_{\text{contr}}(s) = \begin{pmatrix}s & 1 & 0 & 0\\ 0 & s & 1 & 0\end{pmatrix} \Rightarrow c^*_{\text{contr}} = (2, 2).$$

The corresponding I/O models [cf. (3.23b)] are parametrized in terms of three parameters, i.e., $\beta = (\beta_1, \beta_2, \beta_3)$:

$$\Theta^*_{I/O}(s, \beta) = \begin{pmatrix}s(\beta_1 s + \beta_2) & (\beta_1 s + \beta_2) & \beta_3 s^2 & s^2\\ 0 & s & 1 & 0\end{pmatrix}$$
$$\Rightarrow c^*_{I/O} = (2, 3).$$

For the above family of models to be controllable, Condition 9.8 implies $\beta_2 \neq 0$. Finally, the smooth I/O models [cf. (3.23c)] are parametrized in terms of four parameters, i.e., $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$:

$$\Theta^*_{pr}(s, \gamma) = \begin{pmatrix}s(\gamma_1 s + \gamma_2) & (\gamma_1 s + \gamma_2) & 0 & s^2\\ 0 & s(\gamma_3 s + \gamma_4) & s^2 + \gamma_3 s + \gamma_4 & 0\end{pmatrix} \Rightarrow c^*_{pr} = (2, 4).$$

For the models of the above family to be controllable Condition 9.8 implies $\gamma_2\gamma_4 \neq 0$, i.e., $\gamma_2 \neq 0$, and $\gamma_4 \neq 0$. Thus, in the lexicographic ordering:

$$(0,6) < (2,2) < (2,3) < (2,4),$$

which illustrates Proposition 3.24. Furthermore, each one of the three families defined above contains undominated models, and consequently, no inclusion among them is valid.

The data set $\mathbf{D}$ can also be interpreted as containing discrete-time measurements from a 2-input, 2-output discrete-time system. Each experiment starts at time zero, the system being at rest for $t < 0$. The first experiment consists of two measurements:

$$\mathbf{u}(1) = \mathbf{u}(2) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \quad \mathbf{y}(1) = \begin{pmatrix} 0 \\ 4 \end{pmatrix}, \quad \mathbf{y}(2) = \begin{pmatrix} 0 \\ -3 \end{pmatrix};$$

the second experiment consists of four measurements:

$$\mathbf{u}(1) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad \mathbf{u}(2) = \begin{pmatrix} 3 \\ -2 \end{pmatrix}, \quad \mathbf{u}(3)\begin{pmatrix} 0 \\ -3 \end{pmatrix}, \quad \mathbf{u}(4) = \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

$$\mathbf{y}(1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \mathbf{y}(2) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \mathbf{y}(3) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad \mathbf{y}(4) = \begin{pmatrix} 3 \\ 0 \end{pmatrix}.$$

The problem is to find all minimal-complexity controllable, I/O models which fit the above data. According to part $\mathbf{G}$ of Section II, by substituting the forward shift σ^{-1} for s, we obtain the following models of the discrete-time data set given above: the (unique) minimal complexity controllable model $\Theta^*_{contr}(\sigma^{-1})$; the family of all minimal complexity controllable and I/O models $\Theta^*_{I/O}(\sigma^{-1}, \beta)$, $\beta_2 \neq 0$; the family of all minimal complexity causal (nonanticipating) and I/O models given by $\Theta^*_{I/O}(\sigma^{-1}, \beta)$, where $\beta_1 = \beta_2 = 0$; notice that the models in the latter family are *not* controllable. □

Example B: This is an example of rational interpolation. The data are values $Z(\lambda_i)$ if a 2×2 rational matrix Z, at the frequencies: $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = 2$:

$$Z(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad Z(1) = Z(2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Following part $\mathbf{C}$ of Section II, the above data give rise to the following set $\mathbf{D}$ of purely exponential time series $\mathbf{w}_i \in (\mathbf{R}^4)^{\mathbf{R}}$: $\mathbf{w}_1 = \mathbf{p}_1$, $\mathbf{w}_2 = \mathbf{p}_2$, $\mathbf{w}_3 = \mathbf{p}_3 e^t$, $\mathbf{w}_4 = \mathbf{p}_4 e^t$, $\mathbf{w}_5 = \mathbf{p}_5 e^{2t}$, $\mathbf{w}_6 = \mathbf{p}_6 e^{2t}$, where

$$\mathbf{p}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{p}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{p}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix},$$

$$\mathbf{p}_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{p}_5 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{p}_6 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

The purpose of this example is to recursively model $\mathbf{D}$. Making use of the results of Sections VIII and IX, both the MPUM and all C-MCUM's will be computed.

Step 1: The generating system for $\mathbf{w}_1$ is obtained by means of Formula 8.11; its rows are permuted so that the row degrees are ordered in increasing order [row properness is guaranteed from the structure of (8.11)]:

$$\Theta_1^*(s) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ s & 0 & 0 & 0 \end{pmatrix}.$$

By (9.2) $\mathbf{I}_1 = \{1,2,3\}$ and since v_4 is greater than the first three v_i, there is a unique C-MCUM composed of the first three rows of Θ_1^*:

$$\Theta^*_{contr}(s) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Step 2: From (8.6), (8.8) we obtain the second error time series $\mathbf{e}_2 = \mathbf{w}_2$, i.e., $\epsilon_2 = \mathbf{p}_1$. In order to satisfy the condition of corollary (8.21), we need to pre-process Θ_1^*, so that among the entries of the error time series corresponding to rows of Θ_1^* having the same degree, only one is nonzero. This is achieved by premultiplying with the constant matrix P_2, det $P_2 \neq 0$, which amounts to subtracting the third from the first row and leaving the others unchanged. The new Θ_1^*, the resulting error, and the generating system at the second step are thus given by the expressions:

$$\hat{\Theta}_1^*(s) = P_2\Theta_1^*(s) = \begin{pmatrix} 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ s & 0 & 0 & 0 \end{pmatrix} \Rightarrow \hat{\epsilon}_2 = P_2\epsilon_2$$

$$= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \Rightarrow \Theta_2^*(s) = \Gamma_2(s)\hat{\Theta}_1^*(s)$$

$$= \begin{pmatrix} 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & s \\ s & 0 & 0 & 0 \end{pmatrix}.$$

An index l exists and k does not exist. Hence, we are in case b) of (9.12), which means that the third row has to be eliminated; this implies $\mathbf{I}_2 = \{1,2\}$. Since the degree of the third row is greater than that of the second, there is again a unique C-MCUM which consists of the first two rows of Θ_2^*.

$$\Theta^*_{contr,2}(s) = \begin{pmatrix} 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{pmatrix}$$

Step 3: Again, from (8.6), (8.8) we obtain the third error: $\epsilon_3 = \Theta_2^*(1)\mathbf{p}_3 = (-1\,1\,0\,0)'$. Since the rows of Θ_2^* corresponding to the nonzero elements of the error have the same degree, to satisfy (8.21) preprocessing by a constant matrix P_3 is required; it amounts to replacing the first row of Θ_2^* by the sum of the first two rows and leaving the others unchanged. The new Θ_2^*, the corresponding error, and the generating system at the third step are:

$$\hat{\Theta}_2^*(s) = P_3\Theta_2^*(s) = \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & s \\ s & 0 & 0 & 0 \end{pmatrix} \Rightarrow \hat{\epsilon}_3 = P_3\epsilon_3$$

$$= \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \Theta_3^*(s) = \Gamma_3(s)\hat{\Theta}_2^*(s)$$

$$= \begin{pmatrix} -1 & -1 & 1 & 1 \\ -(s-1) & 0 & s-1 & 0 \\ 0 & 0 & 0 & s \\ s & 0 & 0 & 0 \end{pmatrix}.$$

To recover a controllable minimal-complexity model we notice that $r = 2$, and k can be chosen either $k = 3$ or $k = 4$, since that last two rows of Θ_3^* have the same degree and satisfy Condition 9.11d. Hence, case c) of (9.12) applies and we need to make use of the matrix defined by (9.17b), denoted by A_3. A_3 is thus chosen to add the last row to the second row and leave the others unchanged. The new Θ_3^* is

$$\tilde{\Theta}_3^*(s) = A_3\Theta_3^*(s) = \begin{pmatrix} -1 & -1 & 1 & 1 \\ 1 & 0 & s-1 & 0 \\ 0 & 0 & 0 & s \\ s & 0 & 0 & 0 \end{pmatrix}.$$

Thus, $\mathbf{I}_3 = \{1, 2\}$, and a controllable minimal-complexity model is given by the first two rows of $\tilde{\Theta}_3^*$. A parametrization of *all* such controllable models is given by adding to the second row (scalar) multiples of the third and the fourth rows. According to (9.8) the two parameters α_1 and α_2 have to avoid certain hyperplanes:

$$\Theta_{\text{contr},3}^*(s,\alpha) = \begin{pmatrix} -1 & -1 & 1 & 1 \\ 1+\alpha_1 s & 0 & s-1 & \alpha_2 s \end{pmatrix},$$

$$\alpha = (\alpha_2, \alpha_2) \quad \text{and} \quad \alpha_1 \neq -1 \quad \text{or} \quad \alpha_2 \neq 0.$$

Step 4: The fourth error is $\epsilon_4 = \tilde{\Theta}_3^*(1)\mathbf{p}_4 = (0\ 1\ 1\ 1)'$. Since the last three rows of $\tilde{\Theta}_3^*$ have degree equal to one and the corresponding entries of ϵ_4 are all nonzero, to satisfy the condition of (8.21), we pre-process by P_4, in order to have just one of these entries different from zero. This operation amounts to replacing the second row by the difference of the second and the third; the third by the difference of the third and fourth; and the last by the second. In so doing row reducedness is preserved. The new Θ_3^*, the fourth error, and the resulting Θ_4^* are:

$$\hat{\Theta}_3^*(s) = P_4\tilde{\Theta}_3^*(s) = \begin{pmatrix} -1 & -1 & 1 & 1 \\ 1 & 0 & s-1 & -s \\ s & 0 & 0 & -s \\ 1 & 0 & s-1 & 0 \end{pmatrix} \Rightarrow \hat{\epsilon}_4$$

$$= P_4\epsilon_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \Rightarrow \Theta_4^*(s) = \Gamma_4(s)\hat{\Theta}_3^*(s)$$

$$= \begin{pmatrix} -1 & -1 & 1 & 1 \\ 1 & 0 & s-1 & -s \\ s & 0 & 0 & -s \\ s-1 & 0 & (s-1)^2 & 0 \end{pmatrix}.$$

At this step the first two rows are not involved in the update, $\mathbf{I}_4 = \{1, 2\}$, and hence case a) of (9.12) holds. A parametrization of all controllable minimal-complexity models is given by adding a (scalar) multiple of the third row to the second row. By (9.8) this scalar has to avoid the value -1:

$$\Theta_{\text{contr},4}^*(s,\alpha) = \begin{pmatrix} -1 & -1 & 1 & 1 \\ 1+\alpha s & 0 & s-1 & -(\alpha+1)s \end{pmatrix},$$

$$\alpha \neq -1.$$

Step 5: Here, no pre-processing is required. The various quantities are

$$\epsilon_5 = \Theta_4^*(2)p_5 = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \Rightarrow \Gamma_5(s)$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & s-2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \Rightarrow \Theta_5^*(s) = \Gamma_5(s)\Theta_4^*(s)$$

$$= \begin{pmatrix} -1 & -1 & 1 & 1 \\ -(s-2) & 0 & -(s-2)(s-1) & s(s-2) \\ s & 0 & 0 & -s \\ s & 0 & s(s-1) & -s \end{pmatrix}.$$

It turns out that $r = 2$ and $k = 3$; thus case c) of (9.12) holds. A_5, defined by (9.17c), is needed; it amounts to adding the third row to the second and, in order to simplify Θ_5^*, we subtract the third row from the last:

$$\hat{\Theta}_5^*(s) = A_5\Theta_5^*(s)$$

$$= \begin{pmatrix} -1 & -1 & 1 & 1 \\ 2 & 0 & -(s-2)(s-1) & s(s-3) \\ s & 0 & 0 & -s \\ 0 & 0 & s(s-1) & 0 \end{pmatrix}.$$

It follows that $\mathbf{I}_5 = \{1, 2\}$; a parametrization of all minimal-complexity controllable models is obtained by adding to the second row a polynomial multiple of the third row and a scalar multiple of the fourth:

$$\Theta_{\text{contr},5}^*(s,\alpha) = \begin{pmatrix} -1 & -1 & 1 & 1 \\ 2+s(\alpha_1 s+\alpha_2) & 0 & -(s-2)(s-1)+\alpha_3 s(s-1) & s(s-3)-s(\alpha_1 s-\alpha_2) \end{pmatrix}.$$

According to (9.8) the vector parameter $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ has to avoid the hyperplaces given by

$$\alpha_1 + \alpha_2 + 2 \neq 0 \text{ and } \textit{either } 2\alpha_1 + \alpha_2 + 1 \neq 0 \text{ or } \alpha_3 \neq 0.$$

Step 6: Finally, we get:

$$\epsilon_6 = \Theta_5^*(2)\mathbf{p}_6 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \end{pmatrix} \Rightarrow \Gamma_6(s)$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & s-2 \end{pmatrix} \Rightarrow \Theta_6^*(s) = \Gamma_6(s)\hat{\Theta}_5^*(s)$$

$$= \begin{pmatrix} -1 & -1 & 1 & 1 \\ 2 & 0 & -(s-2)(s-1) & s(s-3) \\ s & 0 & 0 & -s \\ 0 & 0 & s(s-1)(s-2) & 0 \end{pmatrix}.$$

The first two rows were not involved in the sixth update and hence case a) of (9.12) holds. $\mathbf{I}_6 = \{1, 2\}$. All controllable minimal-complexity models are obtained by setting $\alpha_3 = 0$ in $\Theta_{contr,5}^*$:

$$\Theta_{\text{contr},6}^*(s,\alpha) = \begin{pmatrix} -1 & -1 & 1 & 1 \\ \alpha_1 s^2 + \alpha_2 s + 2 & 0 & -(s-2)(s-1) & (1-\alpha_1)s^2 - (3+\alpha_2)s \end{pmatrix},$$
$$\alpha_1 + \alpha_2 + 2 \neq 0, \quad 2\alpha_1 + \alpha_2 + 1 \neq 0.$$

Thus, the generating system of $\mathbf{D}$ is equal to the product of the generating systems of the six error time series: $\Theta_6^* = \hat{\Gamma}_6\hat{\Gamma}_5\hat{\Gamma}_4\hat{\Gamma}_3\hat{\Gamma}_2\hat{\Gamma}_1$, where in the notation used above

$$\hat{\Gamma}_1 := \Theta_1^*, \quad \hat{\Gamma}_2 := \Gamma_2 P_2, \quad \hat{\Gamma}_3 := A_3\Gamma_3 P_3, \quad \hat{\Gamma}_4 := \Gamma_4 P_4,$$
$$\hat{\Gamma}_5 := A_5\Gamma_5, \quad \hat{\Gamma}_6 := \Gamma_6.$$

The cascade interconnection of the two-ports defined by the error generating systems as shown in Fig. 3, provides the system theoretic interpretation of the recursive update.

In order to solve the original interpolation problem, we need to consider only controllable systems. According to the definition of $\mathbf{D}$, the first two entries of $\mathbf{w}$ must be the input variables and the other two entries must be the output variables: $\mathbf{w} = (\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{y}_1\ \mathbf{y}_2)'$. It turns out that the 2×2 matrix formed by the last two columns of $\Theta_{contr,6}^*(s, \alpha)$ is always nonsingular; hence all controllable systems are also I/O systems. Finally, it is easy to check that smoothness implies the additional constraint $\alpha_1 \neq 2$. A parametrization of all transfer functions (interpolants) which have minimal McMillan degree equal to 2, is:

$$Z(s) = \frac{1}{(2-\alpha_1)s^2 - (\alpha_2+6)s + 2} \cdot \begin{pmatrix} (s-1)(s-2) & (1-\alpha_1)s^2 - (\alpha_2+3)s \\ (1-\alpha_1)s^2 - (\alpha_2+3)s & (s-1)(s-2) \end{pmatrix},$$

were the parameters α_2, α_2 satisfy the constraints given above. Furthermore, notice that the I/O and smooth I/O families of controllable systems and their complexities, defined by (3.23a–c), are given by

$$\Theta_{pr,6}^*(s,\gamma) = \Theta_{I/O,6}^*(s,\beta) = \Theta_{\text{contr}}, 6^*(s,\alpha)$$

$$\text{where } \gamma = \alpha \text{ with } \alpha_1 \neq 2 \text{ and } \beta = \alpha$$

$$c_{pr,6}^* = c_{I/O,6}^* = c_{\text{contr},6}^* = (2,2).$$

Finally $(1,0) < (2,0) < (2,1) = (2,1) < (2,2) = (2,2)$ are the complexities of the controllable minimal-complexity models at each one of the six steps described above. □

Acknowledgement

The first author wishes to thank Paul Fuhrmann, Siep Weiland, Martin Rickli and Anton Stoorvogel for their helpful comments during the preparation of this manuscript.

References

[1] A. C. Antoulas, "On recursiveness and related topics in linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-31, pp. 1121–1135, 1986.

[2] A. C. Antoulas and B. D. O. Anderson, "On the scalar rational interpolation problem," *IMA J. Math. Contr. Info.*, vol. 3, pp. 61–88, 1986.

[3] ——, "On the stable rational interpolation problem," *Linear Algebra and its Applications*, vols. 122–124, pp. 301–309, 1989.

[4] A. C. Antoulas, "The cascade structure in system theory," in *Three Decades of Mathematical System Theory*, H. Nijmeijer and H. Schumacher Eds., Springer Lecture Notes in Control and Information Sciences, vol. 135, pp. 1–18. New York: Springer-Verlag, 1989.

[5] B. D. O. Anderson and A. C. Antoulas, "Rational interpolation and state variable realizations," *Linear Algebra Applications*, vols. 137/138, pp. 479–509, 1990.

[6] A. C. Antoulas and B. D. O. Anderson, "State space and polynomial approaches to rational interpolation," in *Progress in Systems and Control Theory: Realization and Modeling in System Theory*, M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran, Eds., vol. I, pp. 73–82. Boston, MA: Birkhäuser, 1990.

[7] A. C. Antoulas and J. C. Willems, "Minimal rational interpolation and Prony's method," in *Analysis and Optimization of Systems*, A. Bensoussan and J. L. Lions, Eds., Lecture Notes in Control and Information Sciences, vol. 144, pp. 297–306. New York: Springer-Verlag, 1990.

[8] A. C. Antoulas, J. A. Ball, J. Kang, and J. C. Willems, "On the solution of the minimal rational interpolation problem," *Linear Algebra Applications*, vols. 137/138, pp. 511–573, 1990.

[9] A. C. Antoulas, "On minimal realizations: A polynomial approach," *Syst. Contr. Lett.*, vol. 14, pp. 319–324, 1990.

[10] A. C. Antoulas, T. Matsuo, and Y. Yamamoto, "The realization problem: Linear deterministic systems," in *Mathematical System Theory*, A. C. Antoulas Ed., pp. 191–212. New York: Springer-Verlag, 1991.

[11] A. C. Antoulas, "Recursive modeling of discrete-time time series," *IMA Volume in Linear Algebra for Control Theory*, P. van Dooren and B. F. Wyman, Eds., vol. 62, pp. 1–20, Springer, 1993.

[12] J. A. Ball, I. Gohberg, and L. Rodman, *Interpolation of Rational Matrix Functions*. Boston, MA: Birkhäuser 1990.

[13] O. H. Bosgra, "On parametrization of the minimal partial realization problem," *Syst. Contr. Lett.*, vol. 3, pp. 181–187, 1983.

[14] P. A. Fuhrmann, "Algebraic system theory: An analyst's point of view," *J. Franklin Institute*, vol. 301, pp. 521–540, 1976.

[15] ——, "A matrix Euclidean algorithm and matrix continued fraction expansion," *Syst. Contr. Lett.*, vol. 3, pp. 263–271, 1983.

[16] ——, "Algebraic methods in system theory," in *Mathematical System Theory*, A. C. Antoulas, Ed., pp. 233–265. New York: Springer-Verlag, 1991.

[17] I. Gohberg, M. A. Kaashoek, and L. Lerer, "A directional partial realization problem," *Syst. Contr. Letts.*, vol. 17, pp. 305–314, 1991.

[18] C. Heij, *Deterministic Identification of Dynamical Systems*, Lecture Notes on Control and Information Sciences, vol. 127. New York: Springer-Verlag, 1989.

[19] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

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

[21] R. E. Kalman, P. L. Falb, and M. A. Arbib, *Tropics in Mathematical System Theory*. New York: McGraw-Hill, 1969.

[22] L. Ljung, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[23] J. M. Schumacher, "Transformations of linear systems under external equivalence," *Linear Algebra Applications*, vol. 102, pp. 1–33, 1988.

[24] M. Van Barel and A. Bultheel, "A canonical matrix continued fraction solution of the minimal realization problem," *Linear Algebra Applications*, vol. 122–124, pp. 973–1002, 1990.

[25] J. C. Willems, "From time series to linear system, part I: Finite dimensional, linear, time invariant systems," *Automatica*, vol. 22, pp. 561–580, 1986.

[26] ——, "From time series to linear system, part II: Exact modeling," *Automatica*, vol. 22, pp. 675–694, 1986.

[27] ——, "From time series to linear system, part III: Approximate modeling, *Automatica*, vol. 23, pp. 87–115, 1987.

[28] ——, "Models for dynamics," Dynamics Reported, vol. 2, pp. 171–269, 1989.

[29] ——, "Paradigms and puzzles in the theory of dynamical systems," *IEEE Trans. Automat. Contr.*, vol. 36, pp. 259–294, 1991.

Thanos Antoulas (S'73–M'74–SM'85–F'92) studied at the ETH Zurich, where he obtained the Diploma of Electrical Engineering, the Diploma of Mathematics, and the Ph.D. degree in Mathematics in 1975, 1975, and 1980, respectively.

Since 1982, he has been with the Department of Electrical and Computer Engineering, Rice University, Houston, where he is currently a Professor. He has held the position of Lecturer in Mathematics at the ETH Zurich since 1983.

Dr. Antoulas was elected Fellow of the IEEE for his contributions to the theory of linear modeling, in November 1991. In August 1992, he was awarded, together with R. H. Bishop, the Best Paper Prize of the AIAA (American Institute of Aeronautics and Astronautics), GNC (Guidance Navigation and Control) Section. He has been on the editorial board of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL AND SYSTEMS AND CONTROL LETTERS. He is currently on the editorial board of the *SIAM Journal of Control and Optimization*, and a Guest Editor of *Linear Algebra and its Applications*. His research interests are in the general area of system modeling and control.

Jan C. Willems (S'66–M'68–SM'78–F'80) was born in Bruges, Belgium, in 1939. He studied electromechanical engineering at the University of Gent, and received the M.Sc. degree in electrical engineering from the University of Rhode Island, Kingston and the Ph.D. degree in electrical engineering, from the Massachusetts Institute of Technology, Cambridge, MA, in 1968.

From 1968, he was an Assistant Professor at M.I.T., until 1973, when he was appointed to the present position of Professor of Systems and Control at the Department of Mathematics of the University of Groningen, The Netherlands. His present research interests lie mainly in modeling, identification, control, and issues related to the foundations of system theory.

Dr. Willems is a frequent contributor to the field in the form of journal articles and conference presentations. He is on the Editorial Board of a number of journals and book series and, in particular, he serves as Managing Editor of *Systems and Control Letters* and of the *SIAM Journal on Control and Optimization*.