Applied Mathematics and Computation 145 (2003) 821–851

APPLIED MATHEMATICS AND COMPUTATION

www.elsevier.com/locate/amc

The simulation problem for high order linear differential systems

Tommaso Cotroneo [a,*], Jan C. Willems [b]

[a] *Quantitative Strategy, Merrill Lynch International, 2 King Edward Street, EC1A 1HQ London, UK*

[b] *University of Leuven, ESAT/SCD (SITA), Kasteelpaark Arenberg 10, B-3001 Leuven-Heverlee, Belgium*

Abstract

In this paper we consider linear constant coefficient ordinary differential equations of arbitrary order of the form $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ with R and M given, but otherwise arbitrary, polynomial matrices. To these equations we associate a behavior $\mathfrak{B}$ defined as the set of all vector-valued distributions w, f that solve the equations themselves. We think of f as a given distribution, while w is to be found in such a way that the equations defining the behavior are satisfied. Alongside $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ we also consider initial conditions obtained by imposing the value at time $t = 0$ of a linear combination of the variables w and their derivatives, yielding $S(\frac{d}{dt})w(0) = Ta$ with S a polynomial matrix, T a fixed real matrix, and a a real vector. The three main issues we address are:

(i) *Solvability* of the behavior, meaning the problem of finding conditions that assure that corresponding to a (or any) given vector distribution f, a trajectory w exists such that (w, f) belong to $\mathfrak{B}$. This question is formalized and answered in Section 4.
(ii) The *index* of the behavior, which means investigating how the smoothness of the given distribution f is related to smoothness of the solution w. Section 5 is dedicated to this issue.
(iii) *Compatibility of initial conditions*, in which case the initial conditions $S(\frac{d}{dt})w(0) = Ta$ are considered alongside the behavior $\mathfrak{B}$. We first check that $S(\frac{d}{dt})w(0)$ is well defined. If it is, we provide conditions under which there exists a (unique) distribution

* Corresponding author.
E-mail addresses: tommaso_cotroneo@ml.com (T. Cotroneo), jan.willems@esat.kuleuven.ac.be (J.C. Willems).

0096-3003/$ - see front matter
doi:10.1016/S0096-3003(03)00278-9

w such that w, f belong to $\mathfrak{B}$ and such that $S(\frac{d}{dt})w(0) = Ta$ are also satisfied for a (or any) given a. This problem is addressed in Section 6.

We show how our results generalize to behaviors defined by equations of arbitrary order classical properties of behaviors defined by first order systems such as $E\frac{d}{dt}w + Aw = f$.

Keywords: Behaviors; DAE's; Index

1. Motivation and problem statement

The central object of interest of this paper are systems of linear constant coefficient ordinary differential equations of arbitrary order of the form

$$R\left(\frac{d}{dt}\right)w = M\left(\frac{d}{dt}\right)f \tag{1}$$

with R and M given, but otherwise arbitrary, polynomial matrices. The behavior associated to Eq. (1) is

$$\mathfrak{B} = \left\{(w,f) \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{w}+\mathtt{f}})|R\left(\frac{d}{dt}\right)w = M\left(\frac{d}{dt}\right)f\right\} \tag{2}$$

In the following we think of f as a given distribution, while w is to be found in such a way that $(w,f) \in \mathfrak{B}$.

Alongside (2) we also consider initial conditions obtained by imposing the value at time $t = 0$ of a linear combination of the variables w and their derivatives, yielding

$$S\left(\frac{d}{dt}\right)w(0) = Ta \tag{3}$$

with S a polynomial matrix, T a fixed real matrix, and a a real vector specifying the initial conditions.

A modeling example presented in Section 2 shows that Eq. (1), together with initial conditions of the form (3), are a natural outcome of the procedure of describing a dynamical system and its interaction with the environment. By simulation problem we mean the problem of computing, if it exists, a trajectory of the variables w which satisfies both (1) and (3).

Specifically, the three main issues we address concerning the behavior (2) with initial conditions (3) are:

(i) *Solvability* of the behavior, meaning the problem of finding conditions that assure that corresponding to a (or to any) given vector distribution f, a trajectory w exists such that $(w,f) \in \mathfrak{B}$. This question is formalized and answered in Section 4.

(ii) The *index* of the behavior, equivalently investigating how the smoothness of the given distribution f is related to smoothness of the solution w. Section 5 is dedicated to this issue.
(iii) *Compatibility of initial conditions*, in which case the initial conditions (3) are considered alongside (1). We first check whether $S(\frac{d}{dt})w(0)$ is well defined; this means verifying if, for the given f, there exists at least a trajectory w such that $(w, f) \in \mathfrak{B}$ and such that w is sufficiently differentiable for $S(\frac{d}{dt})w$ to be continuous, so that the value $S(\frac{d}{dt})w(0)$ can be computed. In case $S(\frac{d}{dt})w(0)$ is well defined, we provide conditions under which there exists a (unique) trajectory w such that $(w, f) \in \mathfrak{B}$ and such that Eq. (3) are also satisfied for a (or for any) given a. This problem is addressed in Section 6.

In the course of the discussion we also describe constructive computational algorithms which enable to check all of the above properties and illustrate them on the modeling example presented in Section 2.

The three questions we seek to investigate have been classically studied (see [1,2]) for the special case of behaviors defined by first order differential algebraic equations (DAE's) of the form

$$E\frac{d}{dt}w + Aw = f \tag{4}$$

with initial conditions $w(0) = a$ for some real vector a. The example from Section 2 shows that equations of arbitrary order such as (1) together with initial conditions specified by (3) are a more general and natural way of describing the trajectories of a dynamical systems. At each step in our discussion we will take care of pointing out how our results relate to the classical ones for first order DAE's.

Although our main interest is in linear models, we hint to how our results can be set in the framework of non-linear systems.

Section 3 contains an overview of well established results from behavioral systems theory (see [3,4]) concerning equations of the form

$$D\left(\frac{d}{dt}\right)v = 0 \tag{5}$$

Notice that Eq. (1) can be written in the form 5 by taking $D = [R - M]$ and $v = (w, f)$. It can therefore be expected that studying equations of the general form (5) will provide us with the main instruments we need in order to analyse systems of the form $R(\frac{d}{dt})w = M(\frac{d}{dt})f$.

2. A motivating example

Consider the mechanical system in Fig. 1, which can be viewed as a simplified model of a moving robot. The robot arm is modeled as a rigid body of

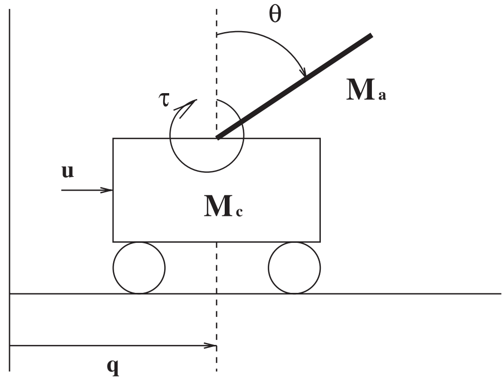

Fig. 1. Cart and arm.

homogeneous mass density with length $2L$ and total mass M_a mounted on a cart with mass M_c on which a force u is exerted. A torque τ is available at the joint between the cart and the arm in order to control the angle θ of the arm.

The variables whose dynamic evolution we wish to model are

$$w = \begin{pmatrix} q \\ \theta \\ u \\ \tau \end{pmatrix}$$

with q the horizontal position of the cart.

The complete model of the above system is, of course, non-linear. However we study its linearized behavior around the equilibrium $u^* = 0$, $\theta^* = 0$, $q^* = 0$, $\tau^* = 0$. The linearized model is given by

$$(M_c + M_a)\frac{d^2}{dt^2}q + M_a L\frac{d^2}{dt^2}\theta = u$$

$$M_a L\frac{d^2}{dt^2}q + \frac{4}{3}M_a L^2\frac{d^2}{dt^2}\theta - M_a L g\theta = \tau$$

with g the gravitational acceleration. By defining the polynomial matrix

$$R_1(\xi) = \begin{pmatrix} (M_c + M_a)\xi^2 & M_a L\xi^2 & -1 & 0 \\ M_a L\xi^2 & \frac{4}{3}M_a L^2\xi^2 - M_a L g & 0 & -1 \end{pmatrix}$$

the above equations can be written in compact notation as $R_1(\frac{d}{dt})w = 0$.

Notice that we end up with two equations in four variables, in other words the system of equations is underspecified, corresponding to the fact that we are describing an "open" physical system interacting with its environment.

Depending on the application one has in mind, there are many possible ways of adding constraints to the above model. We study the situation in which both the force u and the angle θ are imposed, i.e. $u = f_1$, $\theta = f_2$ with f_1, f_2 given functions of time. This corresponds to asking a desired orientation profile for the arm (e.g., constant) while the cart is being pulled by an external force, for example as result of a disturbance (e.g., wind, vibrations), or of a known force (e.g., a locomotive).

The complete model then becomes $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ with

$$R(\xi) = \begin{pmatrix} 2\xi^2 & \xi^2 & -1 & 0 \\ \xi^2 & \frac{4}{3}\xi^2 - g & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad M(\xi) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{6}$$

As far as initial conditions are concerned, we look for trajectories corresponding to a given initial value of the position and speed of the cart. This means looking for trajectories w that satisfy the above differential equations and also $S(\frac{d}{dt})w(0) = a$ for

$$S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \xi & 0 & 0 & 0 \end{pmatrix}$$

and $a \in \mathbb{R}^2$ given.

3. Dynamical systems

The aim of this section is to provide the essential background material from behavioral systems theory which we need in the following, in particular concerning general properties of systems described by Eq. (5). Extensive references on the subject are [3,4]. Throughout we assume the concept and properties of *module* over the polynomial ring $\mathbb{R}[\xi]$ to be known; an extensive reference on the subject is [8]. The notation $\langle R \rangle$ is used to indicate the module over $\mathbb{R}[\xi]$ generated by the columns of the polynomial matrix R.

3.1. Linear differential systems

When modeling a dynamical system one tries to describe how a set of variables of interest evolve as a function of time. We denote these variables as vector v. As extensively argued in behavioral systems theory and also pointed out in the example from Section 2, this typically leads to a set of differential algebraic equations of the form $F(v, \frac{d}{dt}v, \ldots, \frac{d^L}{dt^L}v, t) = 0$ which the vector valued time function $v = (v_1, \ldots, v_v)$ needs to satisfy. If F does not explicitly depend on t, the system is said to be *time-invariant*. Most of our attention will be devoted to the subclass of systems which have this feature and are also *linear*. In this case we deal with a set of constant coefficient linear differential equations

$$D_0 v + D_1 \frac{\mathrm{d}}{\mathrm{d}t} v + \cdots + D_L \frac{\mathrm{d}^L}{\mathrm{d}t^L} v = 0 \tag{7}$$

with $D_i \in \mathbb{R}^{p \times \mathtt{v}}$, $i = 0, 1, \ldots, L$, where p denotes the number of equations. Notice that algebraic constraints are included in the above class of equations. With the polynomial matrix $D(\xi) = D_0 + D_1\xi + \cdots + D_L\xi^L \in \mathbb{R}[\xi]^{p\times q}$, we can write Eq. (7) in the shorthand form (5)

$$D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) v = 0$$

The set of solutions to the above equations represents all admissible trajectories for the v variables, in other words the dynamic behavior of the physical system being modeled. In general we take v a distribution, therefore the behavior $\mathfrak{B}$ specified by (7) is formally defined as:

$$\mathfrak{B} = \left\{ v \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{v}}) | D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) v = 0 \text{ in distributional sense} \right\}$$

where $\mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{v}})$ denotes the set of $\mathbb{R}^{\mathtt{v}}$ valued distributions on $\mathbb{R}$. For obvious reasons, we call (5) a *kernel representation* of $\mathfrak{B}$, and write $\mathfrak{B} = \ker(D(\frac{\mathrm{d}}{\mathrm{d}t}))$.

It turns out that, while D uniquely specifies $\mathfrak{B}$, the converse is not true, in other words different polynomial matrices may induce a kernel representation of the same behavior. In fact, if D, D_1 are both polynomial matrices with $\mathtt{w}$ columns,

$$\ker\left(D\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\right) = \ker\left(D_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\right) \Longleftrightarrow \langle D^{\mathrm{T}} \rangle = \langle D_1^{\mathrm{T}} \rangle$$

In other words, any behavior admits many different kernel representations, but is associated with one and only one submodule of $\mathbb{R}^{\mathtt{v}}[\xi]$, namely the module generated by the transposes of the rows of one (and therefore all) of its possible kernel representations.

Two different kernel representations of a same behavior are called *equivalent*. Among all equivalent representations of $\mathfrak{B} = \ker(D(\frac{\mathrm{d}}{\mathrm{d}t}))$ we call *minimal* those corresponding to polynomial matrices with $c(\mathfrak{B})$ rows, where $c(\mathfrak{B})$ is the minimal number of generators for the submodule $\langle D^{\mathrm{T}} \rangle$. It can be shown that minimal representations correspond to polynomial matrices D which are of full row rank (that is, $D \in \mathbb{R}^{c(\mathfrak{B})\times \mathtt{v}}[\xi]$ has a $c(\mathfrak{B}) \times c(\mathfrak{B})$ submatrix with non-zero determinant) and that if D_1 induces a minimal representation for $\mathfrak{B} = \ker(D(\frac{\mathrm{d}}{\mathrm{d}t}))$ then $D = U\begin{bmatrix} D_1 \\ 0 \end{bmatrix}$ with U an *unimodular* polynomial matrix (that is, a matrix U which admits a polynomial inverse or, equivalently, such that $\det(U)$ is a non-zero constant).

As shown by the example in Section 2, we are typically dealing with under-specified sets of equations, corresponding to the fact that we model open dynamical systems. It can be shown that the difference $\mathtt{v} - c(\mathfrak{B})$ between number of variables and number of rows of a minimal kernel representation of a behavior exactly corresponds to the number of components of v which are left unconstrained. These variables are, for obvious reasons, called *free variables* and can be intuitively interpreted as inputs in the traditional sense. Although the number of free variables is fixed, notice that we leave unspecified which among the variables are actually free; in general, in fact, many possible subsets of the v's with $\mathtt{v} - c(\mathfrak{B})$ elements can play this role, in other words the variables which are inputs to the systems can be chosen in many possible ways. This simple observation is in fact one of the main motivations of the behavioral approach.

3.2. Latent variables

When analysing systems it is often useful to be able to characterise the set of trajectories obtained by projecting a given behavior onto a subset of system variables. We call *manifest variables* those on which we project, while the remaining ones are called *latent variables*.

If we indicate by v_{m} the manifest and by v_{l} the latent variables, then the original behavior $\mathfrak{B}$ consists of all trajectories $v = (v_{\mathrm{m}}, v_\ell)$ that satisfy a set of differential equations of the form:

$$F\left(v_{\mathrm{m}}, \frac{\mathrm{d}}{\mathrm{d}t}v_{\mathrm{m}}, \ldots, \frac{\mathrm{d}^L}{\mathrm{d}t^L}v_{\mathrm{m}}, v_\ell, \frac{\mathrm{d}}{\mathrm{d}t}v_\ell, \ldots, \frac{\mathrm{d}^N}{\mathrm{d}t^N}v_\ell, t\right) = 0$$

In the linear time invariant case these become

$$G_0 v_{\mathrm{m}} + G_1 \frac{\mathrm{d}}{\mathrm{d}t}v_{\mathrm{m}} + \cdots + G_L \frac{\mathrm{d}^L}{\mathrm{d}t^L}v_{\mathrm{m}} = C_0 v_\ell + C_1 \frac{\mathrm{d}}{\mathrm{d}t}v_\ell + \cdots + C_N \frac{\mathrm{d}^N}{\mathrm{d}t^N}v_\ell \qquad (8)$$

for suitable real matrices $G_i \in \mathbb{R}^{p\times \mathtt{v}_{\mathrm{m}}}$ and $C_i \in \mathbb{R}^{p\times \mathtt{v}_1}$ where $\mathtt{v}_{\mathrm{m}}$ and $\mathtt{v}_1$ denote the number of manifest and latent variables respectively. Notice that

$$\mathfrak{B} = \ker\left(\left[G\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) - C\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\right]\right) \subseteq \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{v}}_{\mathrm{m}} + \mathtt{v}_1)$$

The behavior $\mathfrak{B}_{\mathrm{m}}$ obtained by projecting $\mathfrak{B}$ on variables v_{m} is defined as

$$\mathfrak{B}_{\mathrm{m}} = \left\{ v_{\mathrm{m}} \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{v}_{\mathrm{m}}}) | \exists \ell \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{v}_1}) \text{ such that } G\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)v_{\mathrm{m}} = C\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)v_\ell \right\}$$

In the following we often refer to $\mathfrak{B}_m$ as the *manifest* behavior obtained by projecting $\mathfrak{B}$ onto v_m; $\mathfrak{B}$ itself is then referred to as the *full* behavior.

It turns out that $\mathfrak{B}_m$ can be deduced in a very crisp way from the equations defining $\mathfrak{B}$. In order to do so we define the set SYZ(C) of *syzygies* of the rows of a polynomial matrix $C \in \mathbb{R}^{p \times \mathtt{w}}[\xi]$ as:

$$\mathrm{SYZ}(C) = \{n \in \mathbb{R}^p[\xi] : n^{\mathrm{T}} C = 0\}$$

It is easily seen that $\mathrm{SYZ}(C) \subseteq \mathbb{R}^n$ is a module over $\mathbb{R}$; in the algebraic literature it is referred to as the *syzygy module* of the rows of C (see [8]). There holds:

Theorem 1. *Let $\mathfrak{B}_m$ be the manifest behavior obtained by projecting on variables v_m the behavior described by $G(\frac{d}{dt})v_m = C(\frac{d}{dt})v_\ell$. Then*

$$(v_m \in \mathfrak{B}_m) \iff \left((n \in \mathrm{SYZ}(C)) \Rightarrow \left(n^{\mathrm{T}}\left(\frac{d}{dt}\right)G\left(\frac{d}{dt}\right)v_m = 0\right)\right)$$

As a consequence of the module structure of $\mathfrak{H}_C$, the above is equivalent to $(v_m \in \mathfrak{B}) \iff (N^{\mathrm{T}}(\frac{d}{dt})G(\frac{d}{dt})v_m = 0)$ for N any matrix such that $\langle N \rangle = \mathrm{SYZ}(C)$; in other words $\mathfrak{B}_m = \ker(N^{\mathrm{T}}(\frac{d}{dt})G(\frac{d}{dt}))$. The importance of this remark is crucial, showing that if we start with the solution set to a system of linear differential equations and project it on a subset of the variables involved, what we come up with can again be written as solution set of a new system of linear differential equations; therefore Theorem 1 is often referred to as the *elimination theorem*.

Constructive algorithms to build N starting from C are described in the literature (e.g. [7,10]) and will not be repeated here. We assume available a procedure $N = \mathtt{Syzygy}(C)$, that has a polynomial matrix C as input and computes a matrix N such that $\langle N \rangle = \mathrm{SYZ}(C)$ (see e.g. the command `axb` of [12] for a possible implementation). This way, Theorem 1 also provides a constructive way of building a kernel representation of a behavior starting form its specification using latent variables. One can, in fact, easily write a procedure `Elim` that has polynomial matrices G and C as input and outputs a polynomial matrix D such that $D(\frac{d}{dt})v_m = 0$ is the kernel representation of the manifest behavior of $G(\frac{d}{dt})v_m = C(\frac{d}{dt})v_\ell$.

Algorithm 2

$D = \mathtt{Elim}(G, C)$;

$N = \mathtt{Syzygy}(C)$;
$D = N^{\mathrm{T}} C$;

Although the original rationale for introducing latent variables in behavioral systems theory stems from issues related to modeling of interconnected systems (see [3]), we shall see in the sequel that, thanks to Theorem 1 and Algorithm 2, they also provide a very powerful instrument when it comes to manipulation and analysis of linear differential equations and their solutions.

4. Solvability

We now address the solvability of $R(\frac{d}{dt})w = M(\frac{d}{dt})f$, namely establishing conditions under which this differential equation admits a solution w for a given, or for any, f. Although rather standard, these results are a necessary prerequisite to discussing problems of smoothness of solutions and compatibility with initial conditions.

The first question we address is the solvability of the behavior, defined in the following:

Definition 3. Let $R \in \mathbb{R}^{p\times \mathtt{w}}[\xi]$, $M \in \mathbb{R}^{p\times \mathtt{f}}[\xi]$ and $\mathfrak{B} = \{(w,f) \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{w}+\mathtt{f}}) | R(\frac{d}{dt})w = M(\frac{d}{dt})f\}$. $\mathfrak{B}$ is said to be *solvable for given* $f \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{f}})$ if there exists $w \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{w}})$ such that $(w,f) \in \mathfrak{B}$. If this holds for any $f \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{f}})$, then $\mathfrak{B}$ is said to be *solvable for any* f.

We refer to behaviors that are solvable for any f simply as solvable behaviors. In terms of the equations defining the behavior, this means that for any distribution f there exists a distributional solution w to the system $R(\frac{d}{dt})w = M(\frac{d}{dt})f$. It is not difficult to see that this also implies the existence of a classical solution w (i.e., $(R(\frac{d}{dt})w)(t) = (M(\frac{d}{dt})f)(t) \forall t \in \mathbb{R}$) whenever f is taken sufficiently smooth, in particular, when $f \in \mathfrak{C}^{\infty}(\mathbb{R}, \mathbb{R}^{\mathtt{f}})$. This way, our definition of solvability fully covers the one used in the classical literature on first order DAE's (see [1,6,2]) where an equation $F(w, \frac{d}{dt}w, t) = 0$ is defined to be solvable on an open interval $\Upsilon \subseteq \mathbb{R}$ if there exists a w which is continuously differentiable on Υ such that $F(t, w(t), \frac{d}{dt}w(t)) = 0 \ \forall t \in \Upsilon$.

Definition 3 can be recast in the language of latent Variables introduced earlier. If one defines $\mathfrak{B}_{\mathrm{m}}$ as the manifest behavior obtained by projecting $\mathfrak{B}$ on variables f, then solvability for given f is equivalent to asking $f \in \mathfrak{B}_{\mathrm{m}}$, while solvability for any f is equivalent to $\mathfrak{B}_{\mathrm{m}} = \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{f}})$. This way, the conditions for solvability follow as a straightforward corollary to Theorem 1.

Corollary 4. $\mathfrak{B}$ *defined as in* (2) *is solvable for a given* $f \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{f}})$ *if and only if*

$$(n \in \mathrm{SYZ}(R^{\mathrm{T}})) \Rightarrow \left(n^{\mathrm{T}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f = 0\right)$$

It is solvable for any $f \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathrm{f}})$ *if and only if*

$$(n \in \mathrm{SYZ}(R^{\mathrm{T}})) \Rightarrow (n^{\mathrm{T}}M = 0)$$

The condition for solvability for a given f can be interpreted as asking that all linear differential relations which hold for the rows of G should also hold for vector $M(\frac{\mathrm{d}}{\mathrm{d}t})f$, while the condition for solvability for any f asks that the linear relations which hold for the rows of R should also hold for the rows of M. Algorithm 2 can be used to check solvability in terms of the polynomial matrices R and M.

As a special case of the above situation, we obtain that if R is a full row rank polynomial matrix (equivalently, if the only $n \in \mathbb{R}^p[\xi]$ such that $n^{\mathrm{T}}R = 0$ is $n = 0$), then solvability of the behavior defined by (1) is assured for any f. It follows that we can always rewrite a solvable system in an equivalent (i.e. defining the same full behavior) form $R_{\mathrm{r}}(\frac{\mathrm{d}}{\mathrm{d}t})w = M_{\mathrm{r}}(\frac{\mathrm{d}}{\mathrm{d}t})f$ with R_{r} of full row rank. In the following we always consider solvable behaviors; when needed, we also assume, without loss of generality, that the equations defining $\mathfrak{B}$ have R_{r} of full row rank.

This being the standard assumption in most of the paper we wish to sketch a possible way in which R_{r} and M_{r} can be obtained starting from the given R and M.

The algorithm builds a unimodular matrix U such that $UR = \begin{bmatrix} R_{\mathrm{r}} \\ 0 \end{bmatrix}$ with R_{r} of full row rank; the existence of such an U follows from Section 3. As a consequence of $NM = 0$ it follows that $UM = \begin{bmatrix} M_{\mathrm{r}} \\ 0 \end{bmatrix}$; moreover because U is unimodular we know that (w, f) are such that $R(\frac{\mathrm{d}}{\mathrm{d}t})w = M(\frac{\mathrm{d}}{\mathrm{d}t})f$ if and only if $R_{\mathrm{r}}(\frac{\mathrm{d}}{\mathrm{d}t})w = M_{\mathrm{r}}(\frac{\mathrm{d}}{\mathrm{d}t})f$.

In order to build U we use the fact that if N is a matrix Whose columns are a minimal set of generators for $\mathrm{SYZ}(R)$ then N admits a right inverse, in other words, there exists a polynomial matrix L such that $NL = I$ (see e.g. [13] for a proof). If, in turn, N_{r} is a matrix such that $\langle N_{\mathrm{r}}^{\mathrm{T}} \rangle = \mathrm{SYZ}(L)$, then $U = \begin{bmatrix} N_{\mathrm{r}} \\ N \end{bmatrix}$ is unimodular and such that $N_{\mathrm{r}}R = R_{\mathrm{r}}$ is of full row rank; it follows that $M_{\mathrm{r}} = N_{\mathrm{r}}M$.

The following pseudocode procedure called `Reduce` accepts polynomial matrices R and M as input, and outputs polynomial matrices R_{r} and M_{r} computed according to the algorithm just described:

Algorithm 5

```
[R_r, M_R] = Reduce(R, M);
  N = Syzygy(R);
  I = eye(coldim(N)); %I = Identity matrix
  L = Solve(N, I); %L = right inverse of N
  N_r = Syzygy(L); %N_r = Syzygy L
  R_r = N_r * R;
  M_r = N_r * M
```

In the above code, $X = \texttt{Solve}(A, B)$ is a procedure which we assume available to find a solution X to the polynomial equation $AX = B$ for given polynomial matrices A and B. For a possible implementation thereof see, for example, the command `axb` from [12] and the related literature.

5. The index

For a solvable behavior $\mathfrak{B}$, we now wish to investigate how the smoothness of the given forcing function f affects the smoothness of the corresponding solutions w. Establishing this relationship formally leads us to define the *index* of a behavior. This concept also turns out to be crucial in the next section in order to establish what initial conditions are well defined for the given behavior and function f.

Definition 6. Let $\mathfrak{B}$ be defined as in (2) and assume it is solvable according to Definition 3. Define

$$\mathfrak{J} = \{(j_1, \dots, j_{\mathrm{f}}) \in \mathbb{Z}^{\mathrm{m}} | \forall f \text{ with } f_i \in \mathscr{C}^{k+j_i} \exists w \in \mathscr{C}^k \text{ such that } (w, f) \in \mathfrak{B}\}$$

where $\mathscr{C}^k$ for $k < 0$ is the set of all distributions whose $|k|$th integral is a continuous function. Let, moreover, $\preceq$ be the partial ordering on $\mathbb{Z}^{\mathrm{f}}$ defined by $(\alpha_1, \dots, \alpha_{\mathrm{f}}) \preceq (\beta_1, \dots, \beta_{\mathrm{f}}) \iff \alpha_i \leqslant \beta_i \forall i$. If the set J contains a minimal element $\mu = (\mu_1, \dots, \mu_{\mathrm{f}})$ with respect to such a partial ordering we then define μ to be the *multi-index* of $\mathfrak{B}$ and $v = \max_i \mu_i$ to be its *index*.

The multiindex therefore establishes the minimal differentiability requirement on each component of f which assures that a sufficiently differentiable trajectory w can be found; the index provides an upper bound valid for all components of vector f.

The following result shows that $\mathfrak{J}$ always has a minimal element with respect to the partial ordering $\preceq$ so that the multi-index of $\mathfrak{B}$ can always be defined.

Theorem 7. *Let $\mathfrak{B}$ be defined as in* (2) *with $R \in \mathbb{R}^{p\times \mathtt{w}}[\xi]$ a full row rank polynomial matrix. Let P be a square submatrix of R of maximum determinantal degree. Let δ_{ij} be the difference between the degree of the numerator and of the denominator of the (i,j)th entry of the matrix of rational functions $P^{-1}M$. The multi-index of $\mathfrak{B}$ is given by $(\mu_1,\ldots,\mu_{\mathtt{f}})$ with $\mu_j = \max_i \delta_{ij}$.*

Example 8. Consider the behavior described by the first order equations $\frac{\mathrm{d}}{\mathrm{d}t}w + Aw = f$ corresponding, in our notation, to $R(\xi) = \xi I_{\mathtt{w}\times\mathtt{w}} - A$, $M(\xi) = I_{\mathtt{w}\times\mathtt{w}}$. The matrix $(\xi I_{\mathtt{w}\times\mathtt{w}} - A)^{-1}$ is a matrix of proper rational functions. Moreover the difference in degree between the denominator and numerator of all its diagonal entries is equal to 1. It follows that the multi-index is $(-1,\ldots,-1)$ and hence the index is -1. This is in line with the known fact that any $f \in \mathscr{C}^k$ results in a solution $w \in \mathscr{C}^{k+1}$.

Example 9. Consider a system $E\frac{\mathrm{d}}{\mathrm{d}t}w + w = f$ with $E \in \mathbb{R}^{\mathtt{w}\times\mathtt{w}}$ a Jordan block of the form

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

In our notation this corresponds to

$$R(\xi) = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \xi & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \xi & 1 \end{pmatrix}$$

and $M(\xi) = I_{\mathtt{w}\times\mathtt{w}}$.

Since

$$R^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -\xi & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ (-1)^{\mathtt{w}-1}\xi^{\mathtt{w}-1} & (-1)^{\mathtt{w}-2}\xi^{\mathtt{w}-2} & \cdots & -\xi & 1 \end{pmatrix}$$

it follows that the multi-index is $(\mathtt{w}-1,\ldots,0)$ and the hence the index is $\mathtt{w}-1$. This is consistent with the fact that the solution $w = (w_1,\ldots,w_{\mathtt{w}})$ can be explicitly expressed in terms of $f = (f_1,\ldots,f_{\mathtt{w}})$ and its first $(\mathtt{w}-1)$ derivatives as

$$w_j = f_j + \sum_{i=1}^{j-1}(-1)^{j-i}\frac{\mathrm{d}^{j-i}}{\mathrm{d}t^{j-i}}f_i, \quad j = 1,\ldots,\mathtt{w}$$

showing that any f such that $f_i \in \mathscr{C}^{k+\mathtt{w}-i}$ results is $w \in \mathscr{C}^k$.

Example 10. Consider the system $E\frac{\mathrm{d}}{\mathrm{d}t}w + Aw = f$ with

$$R = E\xi + A = \begin{pmatrix} -1 & \xi & 0 & \cdots & 0 & 0 \\ 0 & -1 & \xi & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & \xi \end{pmatrix} \in \mathbb{R}^{(\mathtt{w}-1)\times\mathtt{w}}[\xi]$$

The square minor P of maximal determinantal degree of R corresponds to the last $\mathtt{w}-1$ columns and is such that

$$P^{-1} = \begin{pmatrix} \xi^{-1} & 0 & \cdots & 0 & 0 \\ \xi^{-2} & \xi^{-1} & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \xi^{-(\mathtt{w}-1)} & \xi^{-(\mathtt{w}-2)} & \cdots & \xi^{-2} & \xi^{-1} \end{pmatrix}$$

It follows that the multi-index is $(-1,\dots,-1)$ and thus the index is -1. This coincides with the fact that the equations we are considering are $\frac{\mathrm{d}}{\mathrm{d}t}w_j = w_{j-1} + f_{j-1}, j = 2,\dots,\mathtt{w}$ showing that any f such that $f_i \in \mathscr{C}^{k-1}$ results in $w \in \mathscr{C}^k$ (e.g. take solutions corresponding to $w_1 = 0$).

In the literature on first order DAE's many different concepts of indices have been defined, among which the most relevant appear to be the *differentiation* (see [1,5] and related papers), the *strangeness* (see [2] and related work) and the *perturbation* (see [15]) index. A thorough discussion of these concepts is outside the scope of this chapter. We do recall, however, that the strangeness index s of a time-varying DAE $E(t)\frac{\mathrm{d}}{\mathrm{d}t}w = A(t)w + f(t)$ is defined as the smallest integer such that the given equation can be rewritten in the equivalent form $\frac{\mathrm{d}}{\mathrm{d}t}w_1 = \widehat{A}(t)w_3 + g_1(t)$, $w_2(t) = g_2(t)$ for $w = (w_1, w_2, w_3)$, $\widehat{A}$ a matrix of suitable dimensions, and g a function of the first s derivatives of f. This implies, in particular, that DAE's as those in Examples 8 and 10 have strangeness index $s = 0$, while those in Example 9 have strangeness index $s = \mathtt{w} - 1$. The strangeness index of a generic time-invariant DAE $E\frac{\mathrm{d}}{\mathrm{d}t}w = Aw + f(t)$ with $E\xi + A$ a full row rank polynomial matrix then follows from the three special cases shown in the above examples by transforming the pencil $E\xi + A$ into Kronecker canonical form (see [2,9]). Thus we see that one has $v = s$ whenever $v \geqslant 0$; our definition of index as a smoothness relationship between f and w, however, also allows $v < 0$, corresponding to situations in which $s = 0$. We refer to the literature for further discussion of the relation between differentiation and strangeness indices and for extensions to the non-linear case.

5.1. Computing the index

We can now sketch an algorithm , based on Theorem 7 that allows to compute the multi-index of a system $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ with R a full row rank polynomial matrix. Before doing so, we briefly recall the concept of row reduced (or row proper) form of a polynomial matrix.

Given $R \in \mathbb{R}^{p\times q}[\xi]$, we define its highest row coefficient matrix R^{hc} to be the real matrix whose ith row equals the coefficient vector of the highest power of ξ that occurs in the ith row of R. For example,

$$R = \begin{bmatrix} 3\xi^2 + \xi + 1 & 2 \\ 2\xi & \xi + 1 \end{bmatrix} \Rightarrow R^{hc} = \begin{bmatrix} 3 & 0 \\ 2 & 1 \end{bmatrix}$$

The polynomial matrix R is defined to be *row proper* if the rows of R_{hc} are linearly independent. If R is not row proper, a *row proper form* of R is defined to be any row proper polynomial matrix R_{rp} such that the modules $\langle R\rangle$ and $\langle R_{rp}\rangle$ are the same, and such that R_{rp} is row proper.

It is easily seen that any matrix R actually admits infinitely many row proper forms R_{rp}. Classical algorithms for finding one row proper form R_{rp} starting from R are described in the literature (e.g. in [10]). We assume available a function $R_{rp} = \texttt{rowred}(R)$ whose output R_{rp} is a row-proper form of the input matrix R; a possible implementation thereof is the function $R_{rp} = \texttt{prowred}(R)$ from [12]. Which of the infinitely many row proper forms of R is actually computed depends on the algorithm used. We do not dwell further on this issue because all the considerations we do in the following are valid for any row proper form of R.

We can now show how to compute the multi-index of a system $R(\frac{d}{dt})w = M(\frac{d}{dt})f$; In order to simplify the exposition, we split the algorithm in two subprocedures.

(1) Finding a square minor of R with maximum determinantal degree. We now show how this task can be carried out using a row proper form R_{rp} of R.

Because both R and R_{rp} are of full row rank, there exists a unimodular matrix $U_{rp} \in \mathbb{R}[\xi]^{p\times p}$ such that $U_{rp}R = R_{rp}$, therefore the degrees of the determinants of the $p \times p$ minors of R are equal to those of the corresponding minors of R_{rp}. Moreover, if we let d_i, $i = 1, \dots, p$ be the degree of the ith row of R_{rp} and P any $p \times p$ minor of R_{rp}, then the degree of its determinant is smaller or equal to $d = \sum_i^p d_i$. In particular, if $P_{hc} \in \mathbb{R}^{p\times p}$ is a non-singular minor of the highest coefficient matrix of R_{rp} (such a minor always exists because R_{rp}^{hc} is of full row rank) and P is the corresponding minor of R_{rp}, then the degree of $\det(P)$ is equal to d and therefore maximal. The problem of finding a square minor of R_{rp} with maximum determinantal degree is thus reduced to that of finding a non-singular minor of the real matrix R_{rp}^{hc}; this can be done by standard linear algebra techniques (see e.g. the Matlab command `rank`). In the following we assume available a procedure $m = \texttt{minor}(H)$ that accepts a real matrix H of full row rank as input and outputs a vector m containing the indices of the columns of R that correspond to a non-singular minor of R.

Summarizing what said until now we can sketch a procedure $P = \texttt{maxdet}(R)$ with output P being a square minor of maximal determinantal degree of the full row rank input polynomial matrix R.

Algorithm 11

```
P = maxdet(R)
    R_rp = rowred(R); %R_rp = Row Reduced Form of R
    HC = highdeg(R_rp);
    %HC = Highest coefficient matrix of R_rp
    m = minor(HC);
    P = select(R, m); %Select columns of R
```

In the above code, $\mathrm{HC} = \texttt{highdeg}(R)$ is a procedure that returns the highest coefficient matrix (HC) of the input polynomial matrix R (see `pdegco` from [12] for possible implementation). $P = \texttt{select}(R, m)$, instead, stores in the output matrix P the columns of the input matrix R indexed by the vector m.

(2) Computing the maximum difference of numerator and denominator degree of the entries of each column of a matrix of rational functions of the form $P^{-1}M$. We sketch a procedure that accepts polynomial matrices P and M with P square and non-singular as input and returns a vector of integers d such that the ith component of d corresponds to the maximum difference of numerator and denominator degree in the entries of the ith column of $P^{-1}M$.

Algorithm 12

```
d = maxdiff(P, M)
    P_1 = adjoint(P);
    dp = determinant(P);
    P_2 = P_1 M;
    cd = Degrees of columns of P_2;
    p = Degree of dp
    d = cd-p;
```

What the above procedure effectively does, is write $P^{-1}M$ as $\frac{1}{\det(P)}P_1M$ with P_1 the adjoint of P. If p is the degree of $\det(P)$ and cd is the vector of integers whose ith entry is the degree of the ith column of P_1M, then subtracting p from each entry of d results in a vector whose ith entry is the maximum difference of numerator and denominator degree of the entries of the ith column of $P^{-1}M$.

We refer to the literature (e.g. [14]) for a description of algorithms that compute determinants and adjoints of polynomial matrices; possible implementations thereof are the commands `pdet` and `adj` in the Polynomial Toolbox [12].

The two procedures we have designed can be easily combined to produce an algorithm whose output is the multi-index d of the systems $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ described by the input polynomial matrices R and M.

Algorithm 13

$d = \mathtt{multiindexd}(R, M)$

$P = \mathtt{maxdet}(R)$
$d = \mathtt{maxdiff}(P, M)$

6. Initial conditions

As discussed in the introduction, we now study when, corresponding to a given f, one can find a trajectory w such that (w, f) are in the behavior $\mathfrak{B}$ defined in (1) and such that the initial conditions (3) are satisfied. In this case we say that the initial conditions are compatible with the given $\mathfrak{B}$ and the given f.

The trajectory w is, in general, a distribution. Therefore, before investigating compatibility of (3) with $\mathfrak{B}$ and f, we need to establish conditions under which $S(\frac{d}{dt})w$ is continuous so that asking for its value at a $t = 0$ makes sense. This leads to the following definition of well-posedeness.

Definition 14. Let $\mathfrak{B}$ be defined as in (1). The initial conditions $S(\frac{d}{dt})w(0)$ are said to be *well-posed* for $\mathfrak{B}$ and a given $f \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathrm{f}})$ if there exists a $w \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathrm{w}})$ such that $R(\frac{d}{dt})w = M(\frac{d}{dt})f$, and such that $S(\frac{d}{dt})w \in \mathfrak{C}^0(\mathbb{R}, \mathbb{R}^{\mathrm{s}})$.

The following theorem gives a condition for well-posedeness in terms of the polynomial matrices R and M defining $\mathfrak{B}$, of the polynomial matrix S defining the initial conditions and of the given distribution f.

Theorem 15. *Let* $\mathfrak{B}$ *be defined as in* (1) *and*

$$\mathfrak{B}_s = \left\{ (s, f) \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathrm{s+f}}) | \exists w \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathrm{s+f}}) \text{ such that } (w, f) \in \mathfrak{B} \quad \text{and} \quad S\left(\frac{d}{dt}\right) w = s \right\}$$

Then $S(\frac{d}{dt})w(0)$ *are well posed initial conditions for* $\mathfrak{B}$ *and given* $f \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathrm{f}})$ *if and only if* $f_i \in \mathscr{C}^{\mu_i}$ *with* $(\mu_1, \ldots, \mu_{\mathrm{f}})$ *the multi-index of* B_s.

Notice that, using Theorem 15, well-posedeness can be effectively checked using algorithms for elimination from Section 8 and those for computing the index of the previous section.

We are now ready to address the last issue we planned to tackle, namely establish under what conditions there exists a (unique) solution to $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ which also satisfies $S(\frac{d}{dt})w(0) = Ta$. We start by looking at the existence problem. In order to establish the general result, we first investigate a special case which is of interest in its own right.

Lemma 16. *Let $\mathfrak{B}$ be defined as in* 1 *and assume it is solvable. Let $f \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathrm{f}})$ be such that $w(0)$ are well posed initial conditions for $\mathfrak{B}$. Then there exists a w such that $(w, f) \in \mathfrak{B}$ and $w(0) = a$ if and only if*

$$(n \in \mathbb{R}^{\mathrm{w}},\ b \in \mathbb{R}^{p}[\xi],\ n^{\mathrm{T}} = b^{\mathrm{T}}R) \Rightarrow \left(n^{\mathrm{T}}a = b^{\mathrm{T}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f(0)\right)$$

In the more general case we have:

Theorem 17. *Let $\mathfrak{B}$ be defined as in* (1) *and assume it is solvable. Moreover, let $f \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathrm{f}})$ be such that $S(\frac{d}{dt})w(0)$ are well-posed initial conditions. There exists a w such that $(w, f) \in \mathfrak{B}$ and such that $S(\frac{d}{dt})w(0) = Ta$ if and only if*

$$(n \in \mathbb{R}^{\mathrm{s}},\ b \in \mathbb{R}^{p}[\xi],\ n^{\mathrm{T}}S = b^{\mathrm{T}}R) \Rightarrow \left(n^{\mathrm{T}}Ta = b^{\mathrm{T}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f(0)\right)$$

In particular, if

$$(n \in \mathbb{R}^{s},\ b \in \mathbb{R}^{p}[\xi],\ n^{\mathrm{T}}S = b^{\mathrm{T}}R) \Rightarrow (n^{\mathrm{T}}T = b^{\mathrm{T}}M = 0)$$

then the existence of a w such that $(w, f) \in \mathfrak{B}$ and such that $S(\frac{d}{dt})w(0) = Ta$ is guaranteed for all $a \in \mathbb{R}^{c}$ and all $f \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathrm{f}})$ such that $S(\frac{d}{dt})w(0)$ are well-posed initial conditions.

In order to obtain an intuitive feeling of the above theorem, notice that we are considering conditions that have to hold over the whole real line (the differential equations defining (1)) alongside pointwise constraints (the initial condition (3)). The differential equation, however, also implies pointwise requirements on the solution w; if $b \in \mathbb{R}^{p}[\xi]$ and w is a solution to $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ smooth enough so that $b^{\mathrm{T}}(\frac{d}{dt})R(\frac{d}{dt})w(0)$ is defined, then $b^{\mathrm{T}}(\frac{d}{dt})R(\frac{d}{dt})w(0) = b^{\mathrm{T}}(\frac{d}{dt})M(\frac{d}{dt})f(0)$. On the other hand, $n^{\mathrm{T}}S = b^{\mathrm{T}}R$ implies $n^{\mathrm{T}}S(\frac{d}{dt})w(0) = b^{\mathrm{T}}(\frac{d}{dt}) \times R(\frac{d}{dt})w(0)$. The requirement $n^{\mathrm{T}}Ta = b^{\mathrm{T}}(\frac{d}{dt})M(\frac{d}{dt})f(0)$ can thus be interpreted as asking the initial conditions (3) to be compatible with the pointwise constraints implied by the differential equation. In this sense, it is easily seen that the condition from Theorem 17 is necessary for w to solve both $R(\frac{d}{dt})w = M(\frac{d}{dt})f$

and $S(\frac{d}{dt})w(0) = Ta$. The interesting point is that the condition actually turns out to be also sufficient.

In order to present conditions under which the solution w satisfying given initial conditions is actually unique we need a short intermezzo to show how the concepts of state and state representation of a system are introduced in the behavioral framework we have been using all along (see [11] for more details).

Definition 18. A system $G(\frac{d}{dt})v = C(\frac{d}{dt})x$ with latent variables x and manifest variables w is said to be a *state system* if there exist real matrices E, F, H such that its behavior $\mathfrak{B}$ admits an equivalent representation which is first order in x and of order zero in w, i.e., a representation of the form

$$E\frac{d}{dt}x + Fx + Hw = 0 \tag{9}$$

In this case variables x are called *state variables*. If $\mathfrak{B}_m$ is the manifest behavior obtained by projecting $\mathfrak{B}$ on v, Eq. (9) are called a *state representation* of $\mathfrak{B}$.

If $\mathfrak{B} = \ker(D(\frac{d}{dt}))$, a polynomial matrix X is said to induce a *state map* for $\mathfrak{B}$ if the system with latent variables

$$\begin{aligned} D\left(\frac{d}{dt}\right)w &= 0 \\ X\left(\frac{d}{dt}\right)w &= x \end{aligned} \tag{10}$$

is a state system.

It can be shown that any $\mathfrak{B} = \ker(D(\frac{d}{dt}))$ admits a state map, in other words a set of latent variables having the state property can always be obtained as a function of the manifest variables v.

State representations of a behavior $\mathfrak{B}$ are, of course, not unique. However, the minimal number of state variables associated to $\mathfrak{B}$ is well defined. In other words to any $\mathfrak{B}$ corresponds a number $n(\mathfrak{B})$ such that any state representation of $\mathfrak{B}$ requires at least $n(\mathfrak{B})$ state variables. The number $n(\mathfrak{B})$ is often called the dynamic order, or McMillan degree, of $\mathfrak{B}$. State maps with exactly $n(\mathfrak{B})$ rows are called minimal. We are now able to formulate conditions that the trajectory w corresponding to a given f and to given initial conditions is actually unique.

Theorem 19. *Let* $\mathfrak{B}$ *be defined as in* (1) *and assume it is solvable. Moreover let* $f \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathrm{f}})$ *be such that* $S(\frac{d}{dt})w(0)$ *are well posed initial conditions. There exists a unique* w *such that* $(w, f) \in \mathfrak{B}$ *and such that* $S(\frac{d}{dt})w(0) = Ta$ *if and only if*

1. rank $(R) = \mathtt{w}$,
2. $(n \in \mathbb{R}^s,\ b \in \mathbb{R}[\xi]^p,\ n^{\mathrm{T}}S = b^{\mathrm{T}}R) \Rightarrow (n^{\mathrm{T}}Ta = b^{\mathrm{T}}(\frac{\mathrm{d}}{\mathrm{d}t})M(\frac{\mathrm{d}}{\mathrm{d}t})f(0))$,
3. *S induces a state map for* $\ker(R(\frac{\mathrm{d}}{\mathrm{d}t}))$.

In particular, if conditions (1) *and* (3) *hold together with* $n \in \mathbb{R}^s$, $b \in \mathbb{R}[\xi]^p$, $(n^{\mathrm{T}}S = b^{\mathrm{T}}R) \Rightarrow (n^{\mathrm{T}}T = b^{\mathrm{T}}M = 0)$, *then existence of a unique* w *such that* $(w, f) \in \mathfrak{B}$ *and such that* $S(\frac{\mathrm{d}}{\mathrm{d}t})w(0) = Ta$ *is guaranteed for all* $a \in \mathbb{R}^{c\times 1}$ *and all* $f \in \mathfrak{D}'(\mathbb{R}, \mathbb{R}^{\mathtt{f}})$ *such that* $S(\frac{\mathrm{d}}{\mathrm{d}t})w(0)$ *are well posed initial conditions.*

Comparing the above theorem to Theorem 17, we see that it is condition 2 which assures that a solution exists; conditions 1 and 3, therefore, are those needed to guarantee uniqueness. If w_1 and w_2 were both solutions, their difference $e = w_1 - w_2$ would have to satisfy $R(\frac{\mathrm{d}}{\mathrm{d}t})e = 0$, together with $S(\frac{\mathrm{d}}{\mathrm{d}t})e(0) = 0$. The fact that $\mathrm{rank}(R) = \mathtt{w}$ means that none of the components of e can be chosen freely; while the fact that S defines a state map for $\ker(R(\frac{\mathrm{d}}{\mathrm{d}t}))$ implies that e is a trajectory corresponding to an initial state equal to zero in any state representation of $\ker(R(\frac{\mathrm{d}}{\mathrm{d}t}))$. As shown in the proof of Theorem 19, this is necessary and sufficient to ensure that $e = 0$.

We now illustrate these results by showing that classical results on initial value problems for first order DAE's can be obtained as special cases.

Example 20. Consider the behavior defined by the first order differential equation $E\frac{\mathrm{d}}{\mathrm{d}t}w + Aw = f$, with $\det(E\xi + A) \neq 0$ and initial conditions $w(0) = a$. In our notation this corresponds to $R(\xi) = E\xi + A$, $S(\xi) = T = M(\xi) = I_{\mathtt{w}\times\mathtt{w}}$. The condition $\det(E\xi + A) \neq 0$ assures that the behavior is solvable; its multi-index, instead, depends on the degree of $(E\xi + A)^{-1}$. However, for any f such that w is a continuous function, Lemma 16 tells us that a trajectory w satisfying the given initial condition exists if and only if

$$(n \in \mathbb{R}^{\mathtt{w}},\ b \in \mathbb{R}^{\mathtt{w}}[\xi],\ n^{\mathrm{T}} = b^{\mathrm{T}}E\xi + b^{\mathrm{T}}A) \Rightarrow \left(n^{\mathrm{T}}a = b^{\mathrm{T}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f(0)\right)$$

It is not difficult to see that $b^{\mathrm{T}}E\xi + b^{\mathrm{T}}A$ is a real vector if and only if b^{T} is itself a vector in the left null space of E. If B is any matrix whose rows span this space, then a solution exists if and only if $BAa = Bf(0)$. In particular if E is non-singular (equivalently, if $B = 0$) then a solution exits for arbitrary a and f.

In order to investigate uniqueness of the solution, note that $R(\xi) = E\xi + A$ is a square matrix, and $S = I_{\mathtt{w}\times\mathtt{w}}$ by definition induces a state map for the first order system $E\frac{\mathrm{d}}{\mathrm{d}t}w + Aw = 0$. Conditions 1 and 3 of Theorem 19 are thus satisfied, and we deduce that the solution, if it exists, is also unique.

6.1. Compatibility of initial conditions

We now show algorithms which enable to check the conditions just set forward in Theorems 17 and 19.

Before proceeding we recall that given a polynomial matrix.

$R = R_0 + R_1\xi + \cdots R_L\xi^L \in \mathbb{R}^{p\times \mathtt{w}}[\xi]$, its coefficient matrix $\widetilde{R}$ is defined as the real matrix $\widetilde{R} = [R_0\ R_1 \cdots R_L] \in \mathbb{R}^{p\times L\mathtt{w}}$. In the following we assume available procedures $R = \mathtt{polynomial}(\widetilde{R})$ and $\widetilde{R} = \mathtt{coefficient}(R)$ that transform a polynomial matrix into its coefficient matrix and viceversa (see e.g. commands `ppck` and `punpck` from [12]).

In order to check the condition from Theorem 17 one must notice that given $R \in \mathbb{R}^{p\times \mathtt{w}}[\xi]$ and $S \in \mathbb{R}^{\mathtt{s}\times \mathtt{w}}[\xi]$, the set

$$\mathfrak{R}(R,S) = \{(n,b),\ n \in \mathbb{R}^{\mathtt{s}},\ b \in \mathbb{R}^{p}[\xi],\ |n^{\mathrm{T}}S = b^{\mathrm{T}}R\}$$

is a real vector space. Because $(n,b) \in \mathfrak{R}(R,S)$ implies that the degree of b is smaller or equal to $pL + \lambda$ for λ the degree of S, it also follows that $\mathfrak{R}(R,S)$ is finite dimensional. We are now going to show how to build a real matrix N and a polynomial matrix B such that the columns of $[N\ B]^{\mathrm{T}}$ are a basis for $\mathfrak{R}(R,S)$. Verifying the condition from Theorem 17 is then equivalent to checking $NTa = B(\frac{\mathrm{d}}{\mathrm{d}t})M(\frac{\mathrm{d}}{\mathrm{d}t})f(0)$.

By defining $\delta = pL + \lambda$ and introducing the real vector $\tilde{b}^{\mathrm{T}} = [b_0^{\mathrm{T}}\ b_1^{\mathrm{T}} \cdots b_\delta^{\mathrm{T}}]$ with $b_i \in \mathbb{R}^p$ together with the polynomial matrix

$$R_s = \begin{pmatrix} R \\ \xi R \\ \vdots \\ \xi^{\delta} R \end{pmatrix}$$

one obtains that $(n,b) \in \mathfrak{R}(R,S)$ if and only if $n^{\mathrm{T}}\widetilde{S} = \tilde{b}^{\mathrm{T}}\widetilde{R}_s$. The problem of finding a basis $[N\ B]^{\mathrm{T}}$ for $\mathfrak{R}(R,S)$ is, this way, reduced to that of first finding a basis $[N\ \widetilde{B}]$ for the left null space of the real matrix $\begin{pmatrix} \widetilde{S} \\ -\widetilde{R_s} \end{pmatrix}$, and then recovering matrix B starting from its coefficient matrix $\widetilde{B}$. The first of these two tasks can be performed by standard linear algebra techniques (see e.g. the Matlab comand `null`). The second one is essentially a matter of book keeping which we assume performed by our procedure `polynomial`.

To summarize this discussion we sketch a Matlab procedure called `incond` that builds matrices N and B starting from given polynomial matrices R and S following the steps we presented.

Algorithm 21

$[N, B] = \text{incond } (R, S)$

$\lambda = \text{degree}(S)$; %$\lambda$ = Degree of S
$R_s = \text{shift}(R, \lambda)$; %Build Matrix R_s
$\widetilde{R}_s = \text{coefficient}(R_s)$;
$\widetilde{S} = \text{coefficient}(S)$; %Coefficient matrices s
$K = \text{null}([\widetilde{S}; -\widetilde{R}_s])$; % Basis null space
$N = K[:, 1 : \text{cdim}(S)]$; % Extract N from K
$B = \text{polynomial}(K)$; % Extract B from K

In the above we assumed available an auxiliary procedure `shift` that computes R_s based on the original matrix R and on the degree of S.

In order to check conditions from Theorem 19 we now need to be able to decide if a polynomial matrix S induces a state map for $\ker(R(\frac{d}{dt}))$ for R square and non-singular. In the following we assume available a procedure $X = \texttt{smap}(R)$ that accepts a polynomial matrix R as input, and outputs X, a minimal a state map for $\ker(R(\frac{d}{dt}))$. A detailed description of how this can be achieved is to be found in [11].

As a preliminary remark we recall that there exist unique polynomial matrices Y and V such that

$$S = YR + V \tag{11}$$

with VR^{-1} a matrix of strictly proper rational functions. Algorithms to compute Y and V from R and S are described in [10]; we assume available a procedure $[Y, V] = \texttt{division}(S, R)$ that performs the computation (see `pdiv` from [12] for an implementation).

A result from [11] states that if the polynomial matrix X induces a minimal state map for $\ker(R(\frac{d}{dt}))$, then any polynomial matrix S can be written as $S = WX + YR$ for Y a polynomial matrix, W a real one and WXR^{-1} a matrix of strictly proper rational functions. From 11 it follows that $WX = Y$, therefore the real matrix W can be computed from X and Y using $W = \texttt{Solve}(X, Y)$. Again from [11] it follows that S defines a state map for $\ker(R(\frac{d}{dt}))$ if and only if W is a full column rank matrix a property which can be verified by standard linear algebra techniques (see e.g. the Matlab command `rank`).

We now summarize this discussion in a procedure $ss = \texttt{issmap}(R, S)$ that accepts a square and non-singular polynomial matrix R and a polynomial matrix S as inputs and outputs a boolean variable ss which is true if S defines a state map for $\ker(R(\frac{d}{dt}))$.

Algorithm 22

$ss = \texttt{issmap}(R, S)$

$X = \texttt{smap}\ (R)$; %X = minimal state map for R
$[Y, V] = \texttt{divide}\ (S, R)$; %$S = V + YR,\ VR^{-1}$ strictly proper
$W = \texttt{Solve}\ (X, Y)$; % Solve $Y = WX$
$ss = (\texttt{rank}\ (W) = \texttt{cdim}(W))$ %Compare rank and column dimension of W

6.2. Simulating trajectories

As discussed in the introduction, our final aim is that of simulating (i.e. computing) trajectories of a dynamical systems. The algorithms discussed until now allow to analyze properties of system (1) (solvability, index) together with initial conditions (3) (compatibility). In case the system is solvable and the initial conditions are compatible, one is faced with the task of actually computing a trajectory w that satisfies both (1) and (3). A way of doing so in a numerically reliable way is by introducing state variables that allow to rewrite system (1) as a set of first order ordinary differential equations and initial conditions (3) as a set of constraints on the initial value of the state. In this section we sketch algorithms that perform this rewriting; one can then use classical and very reliable numerical techniques to simulate first order equations with initial conditions (see, e.g. the Matlab command `lsim`).

To keep the discussion more flexible we now discuss in detail the special case in which R is a square and non-singular matrix. The general case can be tackled along the same lines, as shown in the proof of Lemma 16 presented in Section 7.

The key step in the procedure is using $[Z, V] = \texttt{divide}(M, R)$ to compute the unique polynomial matrices Z and V such that $R^{-1}M = Z + R^{-1}V$ with $R^{-1}V$ a matrix of strictly proper rational functions. The system $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ is then equivalent to $R(\frac{d}{dt})e = V(\frac{d}{dt})f$, $e = w - z$ with $z = Z(\frac{d}{dt})f$ (notice that if $(\mu_1, \ldots, \mu_{\mathtt{f}})$ is the multi-index of $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ and f_i is the ith component of f then z will contain derivatives of f_i up to order μ_i, in particular $z = 0$ and $e = w$ if $v < 0$).

Because $R(\frac{d}{dt})e = V(\frac{d}{dt})f$ has index smaller than zero, we can find state variables $x = X_w(\frac{d}{dt})w + X_f(\frac{d}{dt})f$ and real matrices A, G, C that allow us to write it in the equivalent form $\frac{dx}{dt} = Ax + Gfe = Cx$. As a consequence we obtain $\frac{dx}{dt} = Ax + Gfw = Cx + z,\quad z = Z(\frac{d}{dt})f$ as an equivalent representation for $R(\frac{d}{dt})w = M(\frac{d}{dt})f$.

Notice that we have managed to reduce a problem with index greater than zero to one of index smaller or equal to zero, by explicitly computing derivatives of the forcing function f. This feature is shared by the classical algorithms discussed in [6] to transform first order systems of higher differentiation index into systems of differentiation index 0. Reliable numerical or symbolic tools to perform differentiation are, of course, needed for these methods to be effective.

The final step before simulation is transforming the initial conditions.

$S(\frac{d}{dt})w(0) = Ta$ into a set of conditions on the initial value of the state x. This can be done using the fact that from the representation $\frac{dx}{dt} = Ax + Gf$ $w = Cx + z$ it follows that

$$\frac{d^k w}{dt^k} = CA^k x + \sum_{i=0}^{k-1} CA^i G \frac{d^{i-k+1} w}{dt^{i-k+1}} + \frac{d^k z}{dt^k}$$

Using this expression, one can compute a real matrix W and a polynomial matrix Y such that the initial conditions $S(\frac{d}{dt})w(0) = Ta$ are equivalent to $Wx(0) = Ta + Y(\frac{d}{dt})f(0) + S(\frac{d}{dt})z(0)$. We assume available a procedure $[W, Y] = \mathtt{stateincond}(S, A, G, C)$ to perform these computation.

We summarize the discussion of this last section with a pseudocode procedure $[A, G, C, Z, W, Y] = \mathtt{simuldata}(R, M, S)$ that accepts polynomial matrices R, M and S as inputs and outputs real matrices A, G, C and W and polynomial matrices Z, X and Y such that the system $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ with initial conditions $S(\frac{d}{dt})w(0) = Ta$ can be written as $\frac{dx}{dt} = Ax + Gf$ $w = Cx + z$, $z = Z(\frac{d}{dt})f$ with initial conditions on x satisfying $Wx(0) = Ta + Y(\frac{d}{dt})f(0) + S(\frac{d}{dt})z(0)$.

Algorithm 23

$[A, G, C, Z, W, Y] = \mathtt{simuldata}(R, M, S)$

$[Z, V] = \mathtt{divide}(M, R)$; %Reduction to index 0
$X = \mathtt{smap}([R - V])$; %$X$ = Minimal state map for $[R - V_1]$
$[A, G, C] = \mathtt{ssrep}(R, V_1, X)$; %State space representation
$[W, Y] = \mathtt{stateincond}(S, A, G, C)$ %Initial conditions on state

Notice that the above procedure outputs $Z = 0$ if the system $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ has index $v < 0$.

We now return to the example presented in Section 2 and apply the main results of the paper to it. We will take $M_a = M_c = 1$ kg and $L = 1$ m.

Example 24. Matrix R in (6) is square and non-singular (its determinant is $-2\xi^2$), therefore the system is solvable in the sense defined in Section 4. Hence, given any orientation profile for the arm and any force acting on the cart, there exist trajectories for the torque and the position of the cart that satisfy the model equations.

Further,

$$R^{-1}M = \begin{pmatrix} \frac{1}{2\xi^2} & -\frac{1}{2} \\ 0 & 1 \\ 1 & 0 \\ \frac{1}{2} & \frac{5}{6}\xi^2 - g \end{pmatrix}$$

showing that the multi-index of this system is (1,3). Hence the solution w will be continuous if the force exerted on the cart is itself continuous, and the prescribed trajectory θ for the arm is twice continuously differentiable.

As for initial conditions, assume one wants to simulate a trajectory with a given initial value of the torque τ, corresponding to $((\,0 \quad 0 \quad 0 \quad 1\,)w)(0) = \tau_0$. Since $(\,0 \quad 0 \quad 0 \quad 1\,) = b^{\mathrm{T}}R$ for $b^{\mathrm{T}} = (\frac{1}{2} - 1\,\frac{1}{2}\,\frac{5}{6}\xi^2 - g)$, we know from Theorem 17 that this will be possible if and only if $\tau_0 = (\frac{f_1}{2} + \frac{5}{6}\frac{\mathrm{d}^2 f_2}{\mathrm{d}t^2} - gf_2)(0)$.

Assume, instead, we were looking for a given initial value of the position and speed of the cart; equivalently for solutions such that

$$S\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)w(0) = a \quad \text{for } S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \xi & 0 & 0 & 0 \end{pmatrix}$$

for a given $a \in \mathbb{R}^2$. It can be seen that there exist no non-trivial real vector n and polynomial vector b such that $n^{\mathrm{T}}S = b^{\mathrm{T}}R$. Therefore, by Theorem 17, a solution exists for arbitrary initial values a. It can also be seen that S induces a state map (actually, a minimal one) for $\ker(R(\frac{\mathrm{d}}{\mathrm{d}t}))$, thus Theorem 19 assures that such a solution is even unique.

We now wish to simulate the trajectory corresponding to zero initial position and velocity for the cart (i.e. $a = (\,0 \quad 0\,)^{\mathrm{T}}$), to a given sinusoidal force $u = f_1 = 0.1\cos(10t)$ applied on the cart and to a fixed angular position $\theta = f_2 = \frac{\pi}{6}$ for the arm. In order to do so we first apply Algorithm 23 to write the system in first order form. The matrix of rational functions $R^{-1}M$ can be written as $V + Z$ with

$$Z(\xi) = \begin{pmatrix} 0 & -\frac{1}{2} \\ 0 & 1 \\ 1 & 0 \\ \frac{1}{2} & \frac{5}{6}\xi^2 - g \end{pmatrix}$$

a polynomial matrix and

$$V(\xi) = \begin{pmatrix} \frac{1}{2\xi^2} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

a matrix of strictly proper rational functions. It can be shown that a minimal state map for $\ker([R(\frac{d}{dt}) - R(\frac{d}{dt})V(\frac{d}{dt})])$ is induced by

$$X = \begin{pmatrix} \xi & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

yielding a state space representation

$$\frac{d}{dt}x = Ax + Gf, w = Cx + z \quad \text{with } A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

$$G = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad z = Z\left(\frac{d}{dt}\right)f = \begin{pmatrix} -\frac{f_2}{2} \\ f_2 \\ f_1 \\ \frac{f_1}{2} + \frac{5}{6}\frac{d^2 f_2}{dt^2} - gf_2 \end{pmatrix}$$

From this it also follows that the required initial conditions

$$\begin{pmatrix} w(0) \\ \frac{dw}{dt}(0) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

are equivalent to $x(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Fig. 2 shows the trajectories for cart position and torque resulting from the simulation, carried out with Matlab command `lsim`.

7. Proofs

Proof of Theorem 1. $\Rightarrow$) If $v_m \in \mathfrak{B}_m$ there exists a distribution v_ℓ such that $G(\frac{d}{dt})v_m - C(\frac{d}{dt})v_\ell = 0$. Given any $n \in \mathbb{R}^p[\xi]$ this of course implies $n^T(\frac{d}{dt})(G(\frac{d}{dt})w - C(\frac{d}{dt})\ell) = 0$. Therefore $n \in \mathrm{SYZ}(C) \Rightarrow n^T(\frac{d}{dt})G(\frac{d}{dt})w = 0$.

$\Leftarrow$) Let

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

be an unimodular matrix such that

$$UC = \begin{bmatrix} U_1 C \\ U_2 C \end{bmatrix} = \begin{bmatrix} 0 \\ C_2 \end{bmatrix}$$

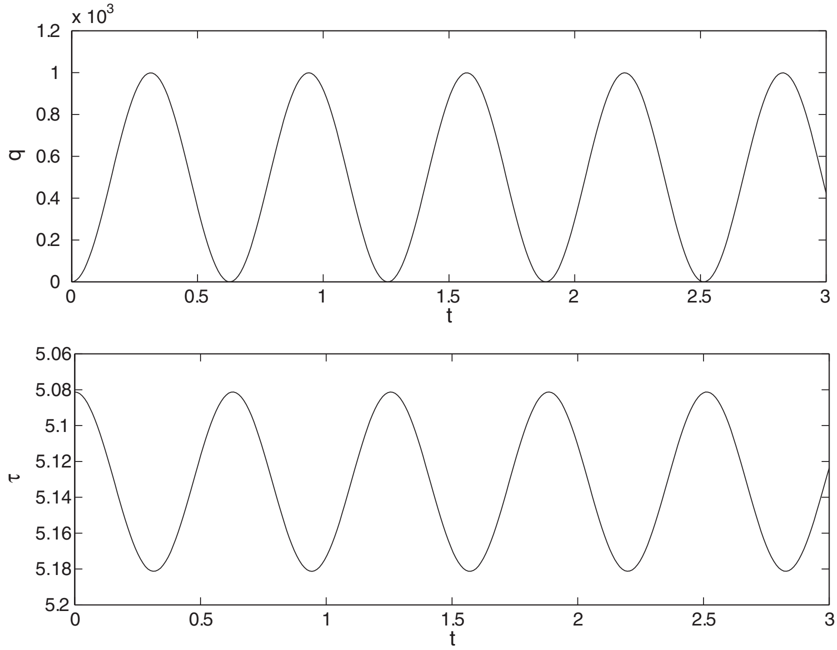

Fig. 2. Cart position and torque.

with $C_2 \in \mathbb{R}^{p' \times 1}[\xi]$ a full row rank polynomial matrix; such a U always exists. Then $\mathfrak{B} = \ker([G(\frac{d}{dt}) - C(\frac{d}{dt})]) = \ker([U(\frac{d}{dt})G(\frac{d}{dt}) - U(\frac{d}{dt})C(\frac{d}{dt})])$. Since C_2 is of full row rank, it can be shown that $C_2(\frac{d}{dt})$ is a subjective map from $\mathfrak{D}'(\mathbb{R}, \mathbb{R}^1)$ to $\mathfrak{D}'(\mathbb{R}, \mathbb{R}^{p'})$, thus the equations $C_2(\frac{d}{dt})v_\ell = G_2(\frac{d}{dt})v_m$ admit a solution v_ℓ for any v_m. The only constraint on v_m is hence given by $G_1(\frac{d}{dt})w = 0$, showing that $\mathfrak{B}_m = \ker(G_1(\frac{d}{dt}))$.

The columns of U_1^T belong to SYZ(C). From the hypothesis $n^T(\frac{d}{dt})G(\frac{d}{dt})w = 0$ for any $n \in$ SYZ(C) it follows $U_1(\frac{d}{dt})G(\frac{d}{dt})w = G_1(\frac{d}{dt})w = 0$, equivalently $w \in \mathfrak{B}_m$ as claimed.

Proof of Theorem 7. We first prove that $(\mu_1, \ldots \mu_f) \in \mathfrak{J}$, meaning that $\forall f$ with $f_i \in \mathscr{C}^{\mu_i+k} \exists w \in \mathscr{C}^k$ such that $R(\frac{d}{dt})w = M(\frac{d}{dt})f$. By linearity it is easily seen that it suffices to prove that $\forall f_i \in \mathscr{C}^{\mu_i+k} \exists w \in \mathscr{C}^k$ such that $R(\frac{d}{dt})w = m_i(\frac{d}{dt})f_i$, with m_i the ith column of M.

Because it is enough to prove it for any k, we choose $k \geqslant 0$ and such that $\mu_i + k \geqslant 0$; this way we only have to work with functions which are at least continuous.

The rational polynomial vector $P^{-1}m_i$ can be written as $P^{-1}m_i = s_i + n_i$ with s_i a strictly proper rational vector and n_i a polynomial one. If $\mu_i \geqslant 0$ then n_i is non-zero and has degree μ_i, otherwise it is zero and the smallest difference in denominator and numerator degree in the elements of s_i is $|\mu_i|$.

If we indicate by y the components of w corresponding to P, by u the remaining ones and by $-Q$ the corresponding columns of R, the equation can be written as $P(\frac{d}{dt})y = Q(\frac{d}{dt})u + P(\frac{d}{dt})s_I(\frac{d}{dt})f_i + P(\frac{d}{dt})n_I(\frac{d}{dt})f_i$, or $P(\frac{d}{dt})y = Q(\frac{d}{dt})u + P(\frac{d}{dt})s_I(\frac{d}{dt})f_i + P(\frac{d}{dt})h_i$ with $h_i = n_I(\frac{d}{dt})f_i \in \mathscr{C}^k$. Because P is full row rank it follows from Theorem 1 and Corollary 4 that u, f_i, h_i can be chosen freely in the equations.

u can thus be chosen freely, therefore also $\mathscr{C}^k$. From Theorem 3.3.13 of [3] and the fact that $P^{-1}Q$, $P^{-1}PS_i$ and $P^{-1}P$ are all proper rational functions, follows that if $\mu_i \geqslant 0$ than y will be as smooth as h_i, therefore $\mathscr{C}^k$. If $\mu_i < 0$ then N_i is zero and, again by the same theorem, it follows that y will be $|\mu_i|$ time smoother than f, thus still $\mathscr{C}^k$.

The same argument shows that if we had taken f such that $f_j = 0$, $j \neq i$ and $f_i \in \mathscr{C}^{k+\mu_i-1}$ then $w \notin \mathscr{C}^k$, thus proving that $(\mu_1, \ldots, \mu_i - 1, \ldots, \mu_{\mathrm{f}})$ does not belong to $\mathfrak{J}$, therefore that $(\mu_1, \ldots, \mu_{\mathrm{f}})$ is the smallest element in $\mathfrak{J}$.

Proof of Theorem 15. By Theorem 1 $\mathfrak{B}_s$ can be described as all (f, s) solving equations $F(\frac{d}{dt})f = L(\frac{d}{dt})s$ for suitable polynomial matrices F and S. These equations are easily seen to be solvable as a consequence of solvability of $R(\frac{d}{dt})w = M(\frac{d}{dt})f$. One can then define the sets

$$\mathfrak{J}_i = \left\{ j \in \mathbb{Z} | \forall h \in \mathscr{C}^{k+j-1} \exists w : R\left(\frac{d}{dt}\right)w = m_i\left(\frac{d}{dt}\right)h \text{ and } S\left(\frac{d}{dt}\right)w \in \mathscr{C}^k \right\}$$

and, applying the definition of manifest behavior, verify that $\mu_i = \min \mathfrak{J}_i$ with $\mu = (\mu_1, \ldots, \mu_{\mathrm{f}})$ the multi-index of $F(\frac{d}{dt})f = L(\frac{d}{dt})s$. Applying Definition 6 and the arguments from Theorem 7 if follows that $f_i \in C^{\mu_i}$ if and only if $s \in \mathbb{C}^0(\mathbb{R}, \mathbb{R}^{\mathrm{s}})$.

Proof of Lemma 16

$\Rightarrow$) Trivial.

$\Leftarrow$) In order to prove this implication we use the concept of state map and state representation introduced in Definition 18.

Because of the smoothness assumption on f we know that a continuous solution w exists; we now need to show that we can find one satisfying $w(0) = a$.

Assume, without loss of generality, that $R = [P - Q]$ with P square and having maximal determinantal degree among all square minors of R. Accordingly, let $w = (y, u)$ and $a = (a_1, a_2)$. Also assume that $M = PV + PZ$ with V a matrix of strictly proper rational functions and Z a polynomial matrix.

This can always be achieved by writing the matrix of rational functions $P^{-1}M$ as the sum of a polynomial matrix Z and a matrix of strictly proper rational functions V. The equation $R(\frac{d}{dt})w = M(\frac{d}{dt})f$ can then be rewritten as

$$P\left(\frac{d}{dt}\right)e = Q\left(\frac{d}{dt}\right)u + P\left(\frac{d}{dt}\right)V\left(\frac{d}{dt}\right)f \tag{12}$$

with $e = y - z$ and $z = Z(\frac{d}{dt})f$.

As discussed in the proof of Theorem 7 the variables u are free in the above equations. Therefore u can be taken a continuous function satisfying the initial constraint $u(0) = a_2$. Now let $[\,X_1 \quad X_2 \quad X_3\,]$ induce a minimal state map for $\ker[\,P(\frac{d}{dt}) \quad -Q(\frac{d}{dt}) \quad -P(\frac{d}{dt}) \quad V(\frac{d}{dt})\,]$. As a consequence of $P^{-1}Q$ and Z being matrices of proper rational functions, it can be shown that a state representation of (12) can be found having the form

$$\frac{d}{dt}x = Ax + Bu + Ef, \quad y = Cx + Du + z \tag{13}$$

with $z = Z(\frac{d}{dt})f$ and with A, B, C, D, E suitable real matrices. Notice that under the given smoothness hypothesis on f, z is a continuous function.

Given any distributions u and f, and any vector x_0 the distribution $x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}(Bu(\tau) + Ef(\tau))d\tau$ is a solution to $\frac{d}{dt}x = Ax + Bu + Ef$. Assume now we choose any continuous u and any f which satisfies the smoothness assumptions of the theorem; then $y = Cx + Du + z$ with $z = Z(\frac{d}{dt})f$ is a continuous function whose value at $t = 0$ is $y(0) = Cx_0 + Du(0) + z(0)$ (notice that x need not be continuous even if f satisfies the required smoothness assumptions; Cx, however, will be continuous).

We now proceed to show that, under the given assumptions, we can always find an x_0 such that $Cx_0 = a_1 - Da_2 + z(0)$ and thus that we can always find a solution which also meets the initial constraint on y. Classical linear algebra tells us that in order to do this we need to show that whenever k is a real vector of suitable dimension such that $k^T C = 0$ then $k^T a_1 - k^T Da_2 - k^T z(0) = 0$.

As a consequence of $[\,X_1 \quad X_2 \quad X_3\,]$ inducing a minimal state map for $\ker[\,P(\frac{d}{dt}) \quad -Q(\frac{d}{dt}) \quad -P(\frac{d}{dt}) \quad V(\frac{d}{dt})\,]$, we can conclude that there exists a polynomial matrix G of suitable dimensions such that $[\,I \quad -D \quad 0\,] = C[\,X_1 \quad X_2 \quad X_3\,] + G[\,P \quad -Q \quad -PV\,]$. Assume now that k is a real vector such that $k^T C = 0$. This implies $k^T GPV = 0$ and $[\,k^T \quad -k^T D\,] = k^T G[\,P \quad -Q\,]$. By the hypothesis of our theorem, with $n^T = [\,k^T \quad -k^T D\,]$ and $b^T = k^T G$, the last statement also implies $k^T a_1 - k^T Da_2 = (k^T G(\frac{d}{dt})M(\frac{d}{dt})f)(0)$. Write, as above, $M = PV + PZ$, use $k^T GPV = 0$ and $k^T = k^T GP$ to obtain $(k^T G(\frac{d}{dt})M(\frac{d}{dt})f)(0) = k^T z(0)$ and thus the statement.

Proof of Theorem 17. $\Rightarrow$) Trivial.

$\Leftarrow$) We are looking for an element $(w, f) \in \ker$

$$R\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) - M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)t$$

with f given and such that $S(\frac{\mathrm{d}}{\mathrm{d}t})w(0) = Ta$. This is equivalent to looking for a

$$(w, f, s) \in \mathfrak{B}_f = \ker \begin{bmatrix} R(\frac{\mathrm{d}}{\mathrm{d}t}) & -M(\frac{\mathrm{d}}{\mathrm{d}t}) & 0 \\ S(\frac{\mathrm{d}}{\mathrm{d}t}) & 0 & -I_{\mathrm{s}\times\mathrm{s}} \end{bmatrix}$$

where f is given and $s(0) = Ta$. As at the end of Section 4, we can consider $\mathfrak{B}_f$ as the full behavior of a latent variable representation with latent variables w and manifest variables (f, s). Let $\mathfrak{B} = \ker[\,F(\frac{\mathrm{d}}{\mathrm{d}t}) \quad -L(\frac{\mathrm{d}}{\mathrm{d}t})\,]$ for suitable matrices $L \in \mathbb{R}^{r\times \mathrm{s}}$ $F \in \mathbb{R}^{r\times \mathrm{f}}$ be the manifest behavior. The problem is to find $(f, s) \in \mathfrak{B}$ with f given and $s(0) = Ta$.

Since $R(\frac{\mathrm{d}}{\mathrm{d}t})w = M(\frac{\mathrm{d}}{\mathrm{d}t})f$ is solvable, so is $L(\frac{\mathrm{d}}{\mathrm{d}t})s = F(\frac{\mathrm{d}}{\mathrm{d}t})f$. Moreover because of the smoothness hypothesis on the given f, a continuous solution s exists. Now apply Lemma 16, and conclude that a continuous solution s satisfying $s(0) = Ta$ exists if and only if $(n \in \mathbb{R}^{\mathrm{s}},\ k \in \mathbb{R}[\xi]^{r},\ n^{\mathrm{T}} = k^{\mathrm{T}}L) \Rightarrow (n^{\mathrm{T}}Ta = k^{\mathrm{T}}(\frac{\mathrm{d}}{\mathrm{d}t})F(\frac{\mathrm{d}}{\mathrm{d}t})f(0))$.

Since $\mathfrak{B}$ is the manifest behavior corresponding to $\mathfrak{B}_f$, the elimination theorem shows that there exist polynomial matrices R_1, M_1, M_2 such that the transpose of the rows of

$$\begin{bmatrix} R & -M & 0 \\ S & 0 & -I \end{bmatrix}$$

generate the same module $\mathfrak{M}$ as the transpose of the rows of

$$\begin{bmatrix} R_1 & M_1 & M_2 \\ 0 & F & -L \end{bmatrix}$$

In particular, given $n \in \mathbb{R}^{\mathrm{s}}$ there exists $k \in \mathbb{R}^{r}[\xi]$ such that $n^{\mathrm{T}} = k^{\mathrm{T}}L$ if and only if $[\,0 \quad k^{\mathrm{T}}F \quad n^{\mathrm{T}}\,]^{\mathrm{T}} \in \mathfrak{M}$ and thus if and only if there exists $b \in \mathbb{R}^{p}[\xi]$ such that $n^{\mathrm{T}}S = b^{\mathrm{T}}R$ and $k^{\mathrm{T}}F = -b^{\mathrm{T}}M$.

The condition $(n \in \mathbb{R}^{\mathrm{s}},\ k \in \mathbb{R}^{r}[\xi],\ n^{\mathrm{T}} = k^{\mathrm{T}}L) \Rightarrow (n^{\mathrm{T}}Ta = k^{\mathrm{T}}(\frac{\mathrm{d}}{\mathrm{d}t})F(\frac{\mathrm{d}}{\mathrm{d}t})f(0))$ is therefore equivalent to $(n \in \mathbb{R}^{\mathrm{s}},\ b \in \mathbb{R}^{p}[\xi],\ n^{\mathrm{T}} = b^{\mathrm{T}}R) \Rightarrow (n^{\mathrm{T}}Ta = b^{\mathrm{T}}(\frac{\mathrm{d}}{\mathrm{d}t})M \times (\frac{\mathrm{d}}{\mathrm{d}t})f(0))$, and the claim is proven.

Proof of Theorem 19. From Theorem 17 we know that condition 2 is necessary and sufficient for existence of a solution. Therefore we now have to prove that conditions 1 and 3 are necessary and sufficient for uniqueness. By linearity this is equivalent to showing that they are necessary and sufficient to ensure that $R(\frac{\mathrm{d}}{\mathrm{d}t})w = 0$ and $S(\frac{\mathrm{d}}{\mathrm{d}t})w(0) = 0$ only admits $w = 0$ as a solution.

$\Rightarrow$) Recall we are working under the assumption that R is a full row rank polynomial matrix. The fact that R needs to be square then follows since otherwise the variables w would contain free components, thus contradicting uniqueness.

Now let $X \in \mathbb{R}^{n\times \mathtt{w}}[\xi]$ induce a minimal state map for $\ker(R(\frac{d}{dt}))$; it can be shown that the corresponding state representation has the form $\frac{d}{dt}x = Ax$, $w = Cx$ for suitable real matrices A and C. Because $x = X(\frac{d}{dt})w$ we conclude that $w = 0 \Rightarrow x = 0$ and, in particular, $x(0) = 0$; vice versa the only solution to $\frac{d}{dt}x = Ax$ satisfying $x(0) = 0$ is $x = 0$, implying $w = 0$. We can thus conclude that $w = 0 \Longleftrightarrow x(0) = 0$.

From the fact that there exists a unimodular matrix U such that

$$\begin{bmatrix} R & 0 \\ X & -I_{n\times n} \end{bmatrix} = U \begin{bmatrix} I_{\mathtt{w}\times\mathtt{w}} & -C \\ 0 & \xi I_{n\times n} - A \end{bmatrix}$$

it also follows that there exists a real matrix W and a polynomial matrix Y such that $S = WX + YR$. The solutions to $R(\frac{d}{dt})w = 0$ satisfying $S(\frac{d}{dt})w(0) = 0$ correspond to those that satisfy $Wx(0) = 0$. Because this equation has to admit $x(0) = 0$ as the unique solution, we conclude that W is injective. From Chapter 6 it follows that S induces a state map for $\ker(R(\frac{d}{dt}))$.

$\Leftarrow$) If G is square and S induces a state map for $\ker(G(\frac{d}{dt}))$ it follows from the definition of state that trajectories which satisfies $G(\frac{d}{dt})w = 0$ and $S(\frac{d}{dt})w(0) = 0$ need to satisfy $\frac{d}{dt}x = Ax$, $w = Cx$, $x(0) = 0$ for $\frac{d}{dt}x = Ax$, $w = Cx$ a state representation of $\ker(G(\frac{d}{dt}))$. It is trivial that the only trajectory for the variables w which satisfies $\frac{d}{dt}x = Ax$, $w = Cx$, $x(0) = 0$ is $w = 0$.

8. Conclusions

In this paper we have discussed the problem of simulating trajectories of dynamical systems. The models we have considered are described by linear differential equations of arbitrary order. The three key issues we have addressed have been those of solvability of the equations, smoothness of solutions (i.e. index) and compatibility of initial conditions. These problems have been classically addressed for first order differential equations and we show how they can be defined in the more general case we are looking at.

The section dedicated to computational algorithms shows how the procedure of analyzing properties of a model and numerically simulating its trajectories can be tackled with help of standard existing software (e.g. Matlab and related toolboxes).

Acknowledgements

The work that resulted in this paper was carried out while the first author was a PhD student at the Rijks Universiteit Groningen. He wishes to thank this institution for providing an excellent working environment.

This research is supported by the Belgian Federal Government under the DWTC program Interuniversity Attraction Poles, Phase V, 2002–2006, Dynamical Systems and Control: Computation, Identification and Modeling.

References

[1] K. Brenan, S. Campbell, L. Petzold, Numerical Solution of Initial-value Problems in Differential–Algebraic Equations, SIAM, 1996.
[2] P. Kunkel, V. Mehrmann, Canonical forms for linear differential-algebraic equations with variable coefficients, Journal of Computational and Applied Mathematics 56 (1994) 225–251.
[3] J.W. Polderman, J.C. Willems, Introduction to Mathematical Systems Theory: A Behavioral Approach, Springer Verlag, 1998.
[4] J.C. Willems, Paradigms and puzzles in the theory of dynamical systems, IEEE Transactions on Automatic Control 36 (1991) 259–294.
[5] K. Brenan, L. Petzold, The numerical solution of higher index differential/algebraic equations by implicit methods, SIAM Journal on Numerical Analysis 26 (4) (1989) 976–996.
[6] C. Gear, L. Petzold, ODE methods for the solution of differential/algebraic systems, SIAM Journal on Numerical Analysis 21 (4) (1984) 716–728.
[7] W. Adams, P. Loustanau, An Introduction to Gröbner Bases, AMS, 1996.
[8] S. Lang, Algebra, third ed., Addison Wesley, 1993.
[9] F.R. Gantmacher, The Theory of Matrices Vol. 2, Chelsea, 1964.
[10] T. Kailath, Linear Systems, Prentice-Hall, 1980.
[11] P. Rapisarda, J.C. Willems, State maps for linear systems, SIAM Journal on Control and Optimization 35 (3) (1997) 1053–1091.
[12] H. Kwakernaak, M. Sebek, Polynomial Toolbox 2. PolyX Ltd., Prague, CZ. http://www.polyx.cz, 1999.
[13] T. Cotroneo, Algorithms in behavioral systems theory, Ph.D. thesis, Department of Mathematics, University of Groningen, 2001. http://www.ub.rug.nl/eldoc/dis/science/t.cotroneo/.
[14] P. Stefanidis, A. Paplinski, M. Gibbard, Numerical operations with polynomial matrices: application to multi-variable compensatro design, Springer Verlag, 1992.
[15] E. Hairer, C. Lubich, M. Roche, The Numerical Solution of Differential-Algebraic Systems by Runge–Kutta Methods, Springer Verlag, 1989.