ELSEVIER

# Modelling dynamical systems using manifest and latent variables

Jan C. Willems*

*Mathematics Institute, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands*

## Abstract

The behavioral approach provides a mathematical language for the modelling of systems, particularly dynamical systems. An introduction to behaviors is given, with emphasis to interconnected systems. This is viewed as consisting of modules, combined with an interconnection architecture. The latter is formalized as a graph with leaves. The elimination theorem is discussed. This allows to obtain behavioral equations involving only manifest variables, starting from models that contain also latent variables. Subsequently, the notions of controllability and observability are cast in this setting. © 2000 IMACS. Published by Elsevier Science B.V. All rights reserved.

*Keywords:* Modelling; Dynamical systems; Manifest variables; Latent variables; Behavior

## 1. Introduction

The purpose of this presentation is to outline the basics of a mathematical language for the modelling, analysis, and the synthesis of dynamical systems. The framework that we will present considers the *behavior* of a system as the main object of study. This paradigm differs in an essential way from the input/output paradigm which has dominated the development of the field of systems and control in the 20th century. This paradigm-shift calls for a reconsideration of many of the basic concepts, of the model classes, of the problem formulations, and of the algorithms in the field.

We will concentrate of a few main themes:
- The basic motivation, in the context of modelling, of the conceptual framework that is used.
- The role of latent variables as they emerge from modelling interconnected systems.
- A discussion of system representations, mainly in the context of systems described by differential equations.
- The notions of controllability and observability in this new setting.
- The formulation of control questions and issues of implementation and design.

* Tel.: +31-50-3633984; fax: +31-50-3633800.
*E-mail address:* j.c.willems@math.rug.nl (J.C. Willems).

This article sketches a mathematical framework that allows to discuss systems in interaction with their environment. However, it is not the purpose to develop mathematical ideas for their own sake. To the contrary, we will downplay mathematical issues throughout. The main aim is to convince the reader that the behavioral framework is a cogent systems theoretic setting that properly deals with physical systems and that approaches modelling as an essential motivation for choosing appropriate concepts.

The behavioral approach is discussed, including the mathematical technicalities, in the recent textbook [1]. A very early reference that contain some of the (immature) ideas is [2]. The three part paper [3–5] provides the first detailed presentation of the behavior framework. It has been further elaborated in [6] and in [7]. This latter reference contains a comprehensive overview. In [8], control is discussed from this perspective. Finally, we mention the article [9] where many of these results are generalized to partial differential equations. Informal expositions of the behavioral approach can be found in [10].

## 2. The behavior

The framework that we use for discussing mathematical models views a model as follows. Assume that we have a phenomenon (i.e. a set of outcomes) that we try to model. Nature (i.e. the reality that governs this phenomenon) can produce certain outcomes. The totality of these possible outcomes (*before* we have modeled the phenomenon) forms a set $\mathbb{U}$, called the *universum*. A *mathematical model* restricts the outcomes that a model declares possible to a subset $\mathcal{B}$ of $\mathbb{U}$; $\mathcal{B}$ is called the *behavior* of the model. We often refer to $(\mathbb{U}, \mathcal{B})$ as a mathematical model.

In the study of dynamical systems we are, more specifically, interested in situations where the outcomes of the phenomena are signals, i.e. maps with independent variables (time, or space, or time and space) and dependent variables (the space where the signals take on their values). In this case the universum is therefore the space of all maps from the set of independent variables to the set of dependent variables. It is hence convenient to distinguish these sets explicitly in the notation: $\mathbb{T}$ (suggesting 'time') for the set of independent, and $\mathbb{W}$ for the set of dependent variables. Whence we define a *dynamical system* as a triple $\Sigma = (\mathbb{W}, \mathbb{T}, \mathcal{B})$ with $\mathcal{B}$, the behavior, a subset of $\mathbb{W}^{\mathbb{T}}$, ($\mathbb{W}^{\mathbb{T}}$ is the standard mathematical notation for the set of all maps from $\mathbb{T}$ to $\mathbb{W}$).

We give a couple of examples. In the first and third $\mathbb{T}$ is time only, while in the second example, Maxwell's equations, $\mathbb{T}$ involves time and space.

1. *Newton's second law* imposes a restriction that relates the position $\vec{q}$ of a point mass and the force $\vec{F}$ acting on it. This relation is $\vec{F} = m\mathrm{d}^2/\mathrm{d}t^2\vec{q}$, with $m$ the mass. This is a dynamical system with $\mathbb{T} = \mathbb{R}, \mathbb{W} = \mathbb{R}^3 \times \mathbb{R}^3$ (typical elements of $\mathcal{B}$ are, $w = (\vec{q}, \vec{F})$: $\mathbb{R} \to \mathbb{R}^3 \times \mathbb{R}^3$) and behavior $\mathcal{B}$ consisting of all maps $t \in \mathbb{R} \mapsto (\vec{q}, \vec{F})(t) \in \mathbb{R}^3 \times \mathbb{R}^3$ that satisfy $\vec{F} = m\mathrm{d}^2/\mathrm{d}t^2\vec{q}$. We will not specify the precise sense of what it means that a function satisfies a differential equation.

2. *Maxwell's equations* provide a typical example of a distributed dynamical system with many independent variables. They describe the possible realizations of the fields $\vec{E}$: $\mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$, $\vec{B}$: $\mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$, $\vec{j}$: $\mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$, and $\rho$: $\mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$. Maxwell's equations are

$$\nabla \cdot \vec{F} = \frac{1}{\varepsilon_0}\rho,$$

$$\nabla \times \vec{F} = -\frac{\partial}{\partial t}\vec{B},$$

$$\nabla \cdot \vec{B} = 0,$$

$$c^2 \nabla \times \vec{B} = \frac{1}{\varepsilon_0} \vec{j} + \frac{\partial}{\partial t} \vec{F},$$

with $\varepsilon_0$ the dielectric constant of the medium and $c^2$ the speed of light in the medium. This defines the system $(\mathbb{R} \times \mathbb{R}^3, \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}, \mathcal{B})$, with $\mathcal{B}$ the set of all fields $(\vec{F}, \vec{B}, \vec{j}, \rho)$: $\mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}$ that satisfy Maxwell's equations.

3. *Kepler's laws* describe the possible motions of the planets in the solar system. This defines a dynamical system with $\mathbb{T} = \mathbb{R}$, $\mathbb{W} = \mathbb{R}^3$, and $\mathcal{B}$ the set of maps $w$: $\mathbb{R} \to \mathbb{R}^3$ that satisfy Kepler's laws: the paths $w$ must be ellipses in $\mathbb{R}^3$ with the sun (assumed in fixed position, say the origin of $\mathbb{R}^3$) in one of the foci; the radius vector from the sun to the planet must sweep out equal areas in equal time, and the ratio of the period of revolution around the ellipse to the major axis must be the same for all $w$'s in $\mathcal{B}$.

These examples fit perfectly our notion of a dynamical system as a triple $\Sigma = (\mathbb{T}, \mathbb{W}, \mathcal{B})$ with $\mathcal{B} \subseteq \mathbb{W}^{\mathbb{T}}$. Of course, the first two examples could be thought of as input/output systems. This already requires some goodwill in the case of Newton's second law in order to avoid a debate of causality in mechanics. But it is inappropriate to force Maxwell's equations (where there are also free variables in the system: the number of equations, eight, being strictly less than the number of variables, 10) into an input/output setting.

First principles laws in physics always state that some outcomes can happen (those satisfying the model equations) while others cannot happen (those violating the model equations). This is a far distance from specifying a system as being driven by free inputs which together with an initial state (whatever that is meant to be) specifies the other variables, the outputs. The behavioral framework treats a model for what it is: an exclusion law.

## 3. Interconnections and latent variables

Systems, especially engineering systems, usually consist of interconnections of subsystems. This feature is crucial in both modelling and design. The aim of this section is to formalize interconnections and to analyze the model structures that emerge from it. We assume throughout finiteness, i.e. we assume that we interconnect a finite number of systems, each with a finite number of terminals, etc.

The building blocks of an interconnected system are *systems with terminals*. Each of these terminals carries variables from a universum, and the laws that governs the system are expressed by a behavior that relates these variables. Formally, a system § with $T$ terminals has a behavior $\mathcal{B} \subset \mathbb{U} = \mathbb{U}_1 \times \mathbb{U}_2 \times \cdots \times \mathbb{U}_T$. If $(u_1, u_2, \ldots, u_T) \in \mathcal{B}$, then we think of $u_k \in \mathbb{U}_k$ as the variables realized at the $k$th terminal.

As an example, consider an electrical component. We view this as an device that can interact with its environment through wires. These wires are the terminals. With each terminal we associate two real variables, the potential $V$ and the current $I$ (agreed to be positive when electrical current flows into the device). The laws of the device specify the behavior which will thus be a subset $\mathcal{B}$ of the universum $\mathbb{R}^2 \times \mathbb{R}^2 \times \cdots \times \mathbb{R}^2 = (\mathbb{R}^2)^T$, where $T$ denotes the number of terminal wires. Usually, the behavior $\mathcal{B}$ will have to satisfy certain restrictions in order for it to qualify as the behavior of an electrical device. For example, *Kirchhoff's current law* and *Kirchhoff's voltage law*. These can be expressed as stating that $((V_1, I_1), \ldots, (V_T, I_T)) \in \mathcal{B}$ must imply $I_1 + I_2 + \cdots + I_T = 0$ and $((V_1 + \alpha, I_1), \ldots, (V_T + \alpha, I_T)) \in \mathcal{B}$ for all $\alpha \in \mathbb{R}$. There may be other requirements, as passivity, etc. but these will not concern us here.

For a thermal terminal, the terminal variables are the heat flow and the temperature. For a mechanical system, the terminal variables are position, and attitude, force and momentum (but it is much more involved to formalize interconnection in this case).

An interconnected system is specified by these subsystems, its building blocks, and by an interconnection architectures. The notion of a *graph with leaves* appears to be the appropriate concept for formalizing an interconnection architecture.

A *graph with leaves* is defined by three (disjoint) sets $(\mathbb{N}, \mathbb{E}, \mathbb{L})$, and two maps $(e, \ell); e: \mathbb{E} \to \tilde{\mathbb{N}}^2$, ($\tilde{\mathbb{N}}^2$ denotes the set of unordered pairs $\{n', n''\}$ with $n', n'' \in \mathbb{N}$) and $\ell: \mathbb{L} \to \mathbb{N}$. The set of $\mathbb{N}$ consists of the *nodes*, $\mathbb{E}$ of the *edges*, $\mathbb{L}$ of the *leaves*; if $e(\alpha) = \{n', n''\}$, then the edge $\alpha$ *connects* the nodes $n'$ and $n''$; if $\ell(\beta) = n$, then the leave $\beta$ is *attached* to the node $n$.

In associating a graph with leaves with an interconnection architecture, the nodes correspond to subsystems with terminals. These are the building blocks that are being connected. Edges that are connected to specific node and leaves that are attached to it, correspond to the terminals of the subsystem in that node. An edge signifies that the corresponding terminal of one subsystem is connected to the corresponding terminal of another (or in the case of a loop, that two terminals of the same system are connected). The leaves signify that the attached terminal is not connected and that it therefore serves as a terminal for the interconnected system.

It is assumed that by interconnecting two terminals by means of an edge, one imposes a restriction on the variables associated with these terminals. For example, if terminal $t_1$ with variables $u_{t_1}$ is connected by an edge with terminal $t_2$ with variables $u_{t_2}$, we assume that a restriction is imposed on the pair $(u_{t_1}, u_{t_2})$. For instance, if $t_1$ and $t_2$ are both electrical terminals, this restriction is $V_{t_1} = V_{t_2}$, $I_{t_1} + I_{t_2} = 0$. If they are thermal terminals, this restriction is $q_{t_1} + q_{t_2} = 0$ (the heat flows are opposite) and $T_{t_1} = T_{t_2}$ (the temperatures are equal). Similar, but more complex, interconnection constraints can be formulated for mechanical connections, etc.

In an interconnection architecture there will usually also be the constraint that edges can only connect terminals that are of the same type (both electrical, both thermal, both mechanical, etc.). Also, a typical system that serves as a building block will have terminals of different type (a motor has electrical and mechanical terminals). However, we do not pursue these ideas here.

The behavior defined by an interconnected system is specified as follows. Its universum equals $\mathbb{U} = \mathbb{U}_{\ell_1} \times \cdots \times \mathbb{U}_{\ell_L}$, where $\mathbb{L} = (\ell_1, \ldots, \ell_L)$ is the set of leaves. The behavior is specified by the behavior of system in the nodes and by the edges. The variables on the terminals connected to a node and the leaves attached to it, must satisfy the laws of the subsystem associated with that node. The variables on the terminals of an edge must satisfy also the interconnection law resulting from the connection.

The resulting behavior $\mathcal{B} \subset \mathbb{U}_{\ell_1} \times \cdots \times \mathbb{U}_{\ell_L}$ of the interconnected system is therefore specified in terms of the behaviors $\mathcal{B}_{n_1}, \ldots, \mathcal{B}_{n_N}$ of the system in the nodes, and the interconnection constraints. The important thing is that the specification of $\mathcal{B}$ involves not only the variables on the leaves, but also those on the edges.

This presence of auxiliary variables in a model is basically an invariant of a first principles modelling procedure: in such a model there will essentially always be auxiliary variables involved in order to specify the laws of the system. It is therefore important to incorporate these auxiliary variables *ab initio* in a modelling framework. This leads to the notion of a model with *manifest* variables (the variables that a model aims at) and *latent* variables (variables that have been introduced in the modelling process). Hence, a *mathematical model with latent variables* is defined as a triple $(\mathbb{U}, \mathbb{L}, \mathcal{B}_{\text{full}})$ with $\mathbb{U}$ the universum of manifest variables, $\mathbb{L}$ the universum of latent variables, and $\mathcal{B}_{\text{full}} \subseteq \mathbb{U} \times \mathbb{L}$ the full behavior. It induces the *manifest systems* $(\mathbb{U}, \mathcal{B})$, with $\mathcal{B} = \{u \in \mathbb{U} | \exists \ell \in \mathbb{L} \text{ such that } (u, \ell) \in \mathcal{B}_{\text{full}}\}$.

A *dynamical system with latent variables* is defined completely analogously as $(\mathbb{T}, \mathbb{W}, \mathbb{L}, \mathcal{B}_{\text{full}})$ with $\mathcal{B}_{\text{full}} \subseteq (\mathbb{W} \times \mathbb{L})^{T}$. The notion of a dynamical system with latent variables is the natural end-point of a modelling process and hence a very natural starting point for the analysis and synthesis of systems. We shall see that latent variables also enter very forcefully in representation questions.

Interconnected systems provide the prime example of the usefulness of behaviors and the inadequacy of input/output thinking. Even if our system, after interconnection, allows for a natural input/output representation, this is unlikely be the case of the subsystem and of the interconnection architecture. If the field of systems and control wants to take modelling seriously, is should retrace the *faux pas* of input/output thinking and cast models in the language of behaviors.

## 4. Differential systems

The 'ideology' that underlies the behavioral approach is the belief that in a model of a dynamical (physical) phenomenon, it is the behavior $\mathcal{B}$, i.e. a set of trajectories $w: \mathbb{T} \to \mathbb{W}$, that is the central object of study. However, as we have seen, in first principles modelling, also latent variables enter *ab initio*. But, the set $\mathcal{B}$ or $\mathcal{B}_{\text{full}}$ of trajectories must be specified somehow, and it is here that differential (and difference) equations enter the scene. Of course, there are important examples where the behavior is specified in other ways (for example, in Kepler's laws for planetary motion), but differential equations are certainly the most prevalent specification of behaviors encountered in applications. For $\mathbb{T} = \mathbb{R}$, and in the case without latent variables, $\mathcal{B}$ then consists of the solutions of a system of differential equations as

$$f_1\left(w, \frac{\mathrm{d}}{\mathrm{d}t}w, \ldots, \frac{\mathrm{d}^N}{\mathrm{d}t^N}w\right) = f_2\left(w, \frac{\mathrm{d}}{\mathrm{d}t}w, \ldots, \frac{\mathrm{d}^N}{\mathrm{d}t^N}w\right).$$

We call these itdifferential systems. In the case of systems with latent variables these differential equation involves both manifest and latent variables. For $\mathbb{T} = \mathbb{R}^n$, this leads to partial differential equations.

Of particular interest (at least in control, signal processing, circuit theory, etc.) are systems with a signal space that is a finite-dimensional vector space and behavior described by linear constant–coefficient differential equations. A *1-D linear time-invariant differential system* is a dynamical system $\Sigma = (\mathbb{R}, \mathbb{W}, \mathcal{B})$, with $\mathbb{W}$ a finite-dimensional (real) vector space, whose behavior consists of the solutions of

$$R\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)w = 0,$$

with $R \in \mathbb{R}^{\bullet \times \bullet}[\xi]$, a real polynomial matrix. We call this a *kernel representation* of the associated linear time-invariant differential system. Of course, the number of columns of $R$ equals the dimension of $\mathbb{W}$. The number of rows of $R$, which represents the number of equations, is arbitrary. In fact, when the row dimension of $R$ is less than its column dimension, as is usually the case, $\mathbb{R}(\mathrm{d}/\mathrm{d}g)w = 0$ is an under-determined system of differential equations which is typical for models in which the influence of the environment is taken into account. The precise definition of what we consider a solution of $R(\mathrm{d}/\mathrm{d}t)w = 0$ is an issue that we will slide over.

The analogue for systems with latent variables, leads to

$$f_1\left(w, \frac{\mathrm{d}}{\mathrm{d}t}w, \ldots, \frac{\mathrm{d}^N}{\mathrm{d}t^N}w, \ell, \frac{\mathrm{d}}{\mathrm{d}t}\ell, \ldots, \frac{\mathrm{d}^N}{\mathrm{d}t^N}\ell\right) = f_2\left(w, \frac{\mathrm{d}}{\mathrm{d}t}w, \ldots, \frac{\mathrm{d}^N}{\mathrm{d}t^N}w, \ell, \frac{\mathrm{d}}{\mathrm{d}t}\ell, \ldots, \frac{\mathrm{d}^N}{\mathrm{d}t^N}\ell\right),$$

relating the (vector of) manifest variables $w$ to the (vector of) latent variables $\ell$. In the linear time-invariant case this becomes

$$R\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)w = M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell,$$

with $R$ and $M$ polynomial. Define the *manifest* behavior of this system as

$$\left\{w\,|\,\exists\ell \text{ such that } R\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)w = M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell\right\}.$$

We call the above differential equation involving $\ell$ a *latent variable* representation of the manifest behavior $\mathcal{B}$. The question occurs whether $\mathcal{B}$ can be described by a linear constant coefficient differential equation. This is the case indeed.

**Theorem 1.** *For any real polynomial matrices $(R, M)$ with $\mathrm{rowdim}(R) = \mathrm{rowdim}(M)$, there exists a real polynomial matrix $R'$ such that the manifest behavior of $R(\mathrm{d}/\mathrm{d}t)w = M(\mathrm{d}/\mathrm{d}t)\ell$ has the kernel representation $R'(\mathrm{d}/\mathrm{d}t)w = 0$.*

The above theorem is called the *elimination theorem*. Its relevance in object-oriented modelling is as follows. As we have seen for the simple electrical circuit discussed in the previous section, a model obtained this way usually involves very many variables and equations, among them many algebraic ones. The elimination theorem tells us that the latent variables may be eliminated, and (in the case of linear time-invariant differential systems) that the number of equations can be reduced to no more than the number of manifest variables. Of course, the order of the differential equation goes up in the elimination process.

## 5. Controllability

An important property in the analysis and synthesis of dynamical systems is controllability. Controllability refers to be ability of transferring a system from one mode of operation to another. By viewing the first mode of operation as undesired and the second one as desirable, the relevance to control and other areas of applications becomes clear. The concept of controllability has originally been introduced in the context of state space systems. The classical definition runs as follows. The system described by the controlled vector-field $(\mathrm{d}/\mathrm{d}t)x = f(x, u)$ is said to be controllable if $\forall a, b, , \exists u$ and $T \geq 0$ such that the solution to $(\mathrm{d}/\mathrm{d}t)x = f(x, u)$ and $x(0) = a$ yields $x(T) = b$. One of the elementary results of system theory states that the finite-dimensional linear system $(\mathrm{d}/\mathrm{d}t)x = Ax + Bu$ is controllable if and only if the matrix $[B\ \ AB\ \ A^2B\cdots A^{\dim(x)-1}B]$ has full row rank. Various generalizations of this result to time-varying, to nonlinear (involving Lie brackets), and to infinite-dimensional systems exist.

A disadvantage of the notion of controllability as formulated above is that it refers to a particular representation of a system, notably a state space representation. Thus a system may be uncontrollable either for the intrinsic reason that the control has insufficient influence on the system variables, or because the state has been chosen in an inefficient way. It is clearly not desirable to confuse these reasons. In the context of behavioral systems, a definition of controllability has been put forward that involves the system variables directly.

Let $\Sigma = (\mathbb{T}, \mathbb{W}, \mathcal{B})$ be a dynamical system with $\mathbb{T} = \mathbb{R}$ or $\mathbb{Z}$, and assume that is time-invariant, that is $\sigma^t \mathcal{B} = \mathcal{B}$ for all $t \in \mathbb{T}$, where $\sigma^t$ denotes the $t$-shift (defined by $(\sigma^t f)(t') = f(t' + t)$); $\Sigma$ is said to

be *controllable* if for all $w_1, w_2 \in \mathcal{B}$ there exists $T \in \mathbb{T}$, $T \geq 0$ and $w \in \mathcal{B}$ such that $w(t) = w_1(t)$ for $t < 0$ and $w(t) = w_2(t - T)$ for $t \geq T$. Thus controllability refers to the ability to switch from any one trajectory in the behavior to any other one, allowing some time-delay.

Two questions that occur are the following: what conditions on the parameters of a system representation imply controllability? Do controllable systems admit a particular representation in which controllability becomes apparent? For linear time-invariant differential systems, these questions are answered in the following theorem.

**Theorem 2.** *Let* $\Sigma = (\mathbb{R}, \mathbb{R}^w, \mathcal{B})$ *be a linear time-invariant differential system. The following are equivalent:*
1. $\Sigma$ *is controllable;*
2. *The polynomial matrix* $R$ *in a kernel representation* $R(\mathrm{d}/\mathrm{d}t)w = 0$ *of* $\mathcal{B}$ *satisfies* rank $(R(\lambda)) = \mathrm{rank}(R)$ *for all* $\lambda \in \mathbb{C}$;
3. *The behavior* $\mathcal{B}$ *is the image of a linear constant–coefficient differential operator, that is, there exists a polynomial matrix* $M \in \mathbb{R}^{w \times \bullet}[\xi]$ *such that* $\mathcal{B} = \{w | w = M(\mathrm{d}/\mathrm{d}t)\ell$ *for some* $\ell\}$.

There exist various algorithms for verifying controllability of a system starting from the coefficients of the polynomial matrix $R$ in a kernel (or a latent variable) representation of $\Sigma$, but we will not enter into these algorithmic aspects.

A point of the above theorem that is worth emphasizing is that, as stated in the above theorem, controllable systems admit a representation as the manifest behavior of the latent variable system of the special form

$$w = M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell.$$

We call this an *image* representation. It follows from the elimination theorem that every system in image representation can be brought in kernel representation. But not every system in kernel representation can be brought in image representation: it is precisely the controllable ones for which this is possible.

The controllability issue has been pursued for many other classes of systems. In particular (more difficult to prove) generalizations have been derived for differential-delay [12,14], for nonlinear, for *n*-D systems [11,13], and, as we will discuss soon, for PDE's. Systems in an image representation have received much attention recently for nonlinear differential–algebraic systems, where they are referred to as *flat* systems [15]. Flatness implies controllability, but the exact relation remains to be studied.

The controllability issue has been pursued for many other classes of systems. In particular (more difficult to prove) generalizations have been derived for differential-delay, nonlinear, and *n*-D systems, and, as we will discuss soon, for PDE's. Systems in an image representation have received much attention recently for nonlinear differential–algebraic systems, where they are referred to as *flat* systems. Flatness implies controllability, but the exact relation remains to be studied.

## 6. Observability

The notion of observability is always introduced hand in hand with controllability. In the context of the input/state/output system $(\mathrm{d}/\mathrm{d}t)x = f(x, u)$, $y = h(x, u)$, it refers to the possibility of deducing,

using the laws of the system, the state from observation of the input and the output. The definition that is used in the behavioral context is more general in that the variables that are observed and the variables that need to be deduced are kept general.

Let $\Sigma = (\mathbb{T}, \mathbb{W}, \mathcal{B})$ be a dynamical system, and assume that $\mathbb{W}$ is a product space: $\mathbb{W} = \mathbb{W}_1 \times \mathbb{W}_2$. Then $w_1$ is said to be *observable* from $w_2$ in $\Sigma$ if $(w_1, w'_2) \in \mathcal{B}$ and $(w_1, w''_2) \in \mathcal{B}$ imply $w'_2 = w''_2$. Observability thus refers to the possibility of deducing the trajectory $w_1$ from observation of $w_2$ and from the laws of the system ($\mathcal{B}$ is assumed to be known).

The theory of observability runs parallel to that of controllability. We mention only the result that for linear time-invariant differential systems, $w_1$ is observable from $w_2$ if and only if there exists a set of differential equations satisfied by the behavior of the system of the following form that puts observability into evidence: $w_1 = R'_2(\mathrm{d}/\mathrm{d}t)w_2$.

## 7. Distributed systems

We now explain the generalization of some of the above concepts and results to constant–coefficient PDE's. Define a *distributed differential system* as an $n$-D system $\Sigma = (\mathbb{R}^n, \mathbb{R}^w, \mathcal{B})$, with behavior $\mathcal{B}$ consisting of the solution set of a system of partial differential equations:

$$R\left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right) w = 0,$$

viewed as an equation in the functions

$$(x_1, \ldots, x_n) = x \in \mathbb{R}^n \mapsto (w_1(x), \ldots, w_w(x)) = w(x) \in \mathbb{R}^w.$$

Here, $R \in \mathbb{R}^{\bullet \times w}[\xi_1, \ldots, \xi_n]$ is a matrix of polynomials in $\mathbb{R}[\xi_1, \ldots, \xi_n]$. Important properties of these systems are their *linearity* (meaning that $\mathcal{B}$ is a linear subspace of $(\mathbb{R}^w)^{\mathbb{R}^n}$), and *shift-invariance* (meaning $\sigma^x \mathcal{B} = \mathcal{B}$ for all $x \in \mathbb{R}^n$, where $\sigma^x$ denotes the $x$-shift, defined by $(\sigma^x f)(x') = f(x' + x)$). We call the above PDE a *kernel representation* of this $n$-D system.

For distributed differential systems with latent variables, this leads to equations of the form

$$R\left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right) w = M\left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right) \ell,$$

with $R$ and $M$ matrices of polynomials in $\mathbb{R}[\xi_1, \ldots, \xi_n]$. This equation relates the (vector of) manifest variables $w$ to the (vector of) latent variables $\ell$. Define the *full behavior* of this system as

$$\mathcal{B}_{\text{full}} = \{(w, \ell)| \text{ the PDE } in(w, \ell) \text{ holds}\}$$

and the *manifest behavior* as

$$\mathcal{B} = \{w| \exists \ell \text{ such that}(w, \ell) \in \mathcal{B}_{\text{full}}\}$$

We call the PDE with latent variables a *latent variable* representation of $\mathcal{B}$. The question again occurs whether $\mathcal{B}$ can itself be described by a set of PDE's. This is the case indeed.

**Theorem 3.** *For any pair of real matrices of polynomials $(R, M)$ in $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_n]$ with $\operatorname{rowdim}(R) = \operatorname{rowdim}(M)$, there exists a real matrix of polynomials $R'$ in $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_n]$ such that the manifest behavior of $\mathcal{B}$ has kernel representation $R'(\partial/\partial x_1, \ldots, \partial/\partial x_n)w = 0$.*

As an illustration of the elimination theorem, consider the elimination of $\vec{B}$ and $\rho$ from Maxwell's equations. The following equations describe the possible realizations of the fields $\vec{F}$ and $\vec{j}$:

$$\varepsilon_0 \frac{\partial}{\partial t} \nabla \cdot \vec{F} + \nabla \cdot \vec{j} = 0, \quad \varepsilon_0 \frac{\partial^2}{\partial t^2} \vec{F} + \varepsilon_0 c^2 \nabla \times \nabla \times \vec{F} + \frac{\partial}{\partial t} \vec{j} = 0.$$

Note that it follows from the elimination theorem that the manifest behavior of a system in image representation, i.e. a latent variable system of the special form

$$w = M \left( \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right) \ell \tag{1}$$

can be described as the solution set of a system of constant coefficient PDE's. Whence, every image of a constant coefficient linear partial differential operator is the kernel of a constant coefficient linear partial differential operator. However, not every kernel of a constant coefficient linear partial differential operator is the image of a constant coefficient linear partial differential operator. The following theorem, obtained in [9], shows that it are precisely the controllable systems that admit an image representation.

**Theorem 4.** *The following statements are equivalent for systems described by constant coefficient linear PDE's:*
1. *$\mathcal{B}$ defines a controllable system,*
2. *$\mathcal{B}$ admits an image representation,*
3. *The trajectories of compact support are dense in $\mathcal{B}$.*

It can be shown that Maxwell's equations define a controllable distributed differential system. Note that an image representation corresponds to what in mathematical physics is called a *potential function* with $\ell$ the potential and $M(\partial/\partial x_1, \ldots, \partial/\partial x_n)$, the partial differential operator that generates elements of the behavior from the potential. An interesting aspect of the above theorem therefore is the fact that it identifies the existence of a potential function with the system theoretic property of controllability and concatenability of trajectories in the behavior. In the case of Maxwell's equations, an image representation is given by

$$\vec{F} = -\frac{\partial}{\partial t} \vec{A} - \nabla \phi,$$

$$\vec{B} = \nabla \times \vec{A},$$

$$\vec{j} = \varepsilon_0 \frac{\partial^2}{\partial t^2} \vec{A} - \varepsilon_0 c^2 \nabla^2 \vec{A},$$

$$\rho = \frac{\varepsilon_0}{c^2} \frac{\partial^2}{\partial t^2} \phi - \varepsilon_0 \nabla^2 \phi,$$

where $\phi \colon \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$ is a scalar, and $\vec{A} \colon \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$ a vector potential. Note that Maxwell's equations consist of eight equations in 10 variables. It turns out that the number of free variables is three. In the

above image representation there are four free latent variables. This can actually be reduced to three, say by putting one component of $\vec{A}$ to zero. A more elegant way of reducing the freedom in the latent variables is by imposing a *gauge*, for example, restricting $\vec{A}$ and $\phi$ to satisfy $c^2 \nabla \cdot \vec{A} + \partial/\partial t \phi = 0$. Imposing this gauge retains the symmetry, but the resulting set of equations yields a latent variable representation of the behavior, not an image representation.

For distributed differential systems, $w_1$ is observable from $w_2$ if and only if there exists a set of annihilators of the behavior of the following form that puts observability into evidence: $w_1 = R_2'(\partial/\partial x_1, \ldots, \partial/\partial x_n)w_2$, with $R_2' \in \mathbb{R}^{\dim(w_1) \times \dim(w_2)}[\xi_1, \ldots, \xi_n]$. We call a latent variable representation of the manifest behavior *observable* if $\ell$ is observable from $w$ in its full behavior. We call it *weakly observable*, if to every $w \in \mathcal{B}$ of compact support, there corresponds a unique $\ell$ that is also of compact support.

For 1-D systems it is easy to show that every controllable linear time-invariant differential behavior $\mathcal{B}$ admits an observable image representation. This, however, does not hold for $n$-D systems, and hence the representation of controllable systems in image representation (i.e. with potential functions) may require the introduction of latent variables that are 'hidden', in the sense that $M(\partial/\partial x_1, \ldots, \partial/\partial x_n)\ell = 0$ has solutions $\ell \neq 0$. This means that however one represents a controllable behavior $\mathcal{B}$ of a PDE as $w = M(\partial/\partial x_1, \ldots, \partial/\partial x_n)\ell$, there may not exist an $N \in \mathbb{R}^{w \times \bullet}[\xi_1, \ldots, \xi_n]$ such that $w = M(\partial/\partial x_1, \ldots, \partial/\partial x_n)\ell$ implies $\ell = N(\partial/\partial x_1, \ldots, \partial/\partial x_n)w$. The latent variables do not be recoverable from the manifest ones by a 'local' differential operator. However, locally observable image representations always exist.

For example, the image representation of the behavior defined by Maxwell's equations in terms of the vector potential $\vec{A}$ field and the scalar potential $\phi$, is not observable (neither is the latent variable representation obtained after imposing the gauge, but then the resulting latent variable representation is weakly observable). In fact, Maxwell's equations are an example of a controllable system that does not allow an observable image representation.

## 8. Conclusions

In this paper, we have covered some highlights of the behavioral approach to systems and control. We view a mathematical model as a subset of an universum. However, in engineering applications, models are invariably obtained by interconnecting subsystems. This leads to the presence in mathematical models of manifest variables (the variables whose behavior the model aims at) and latent variables (the auxiliary variables introduced in the modelling process). Thus the central object in systems theory is a dynamical system with latent variables.

Various problems occur in this framework. For example, the elimination problem: obtaining differential equations for the manifest behavior that contain only the manifest variables. Further, the state space representation problem: obtaining a special latent variable representation in which the latent variables capture the memory of a system. There are many other representation questions, related to image representations, to input/output representations, etc.

In the behavioral framework, the concept of controllability becomes an intrinsic systems property related to concatenability of system trajectories. In the context of latent variable systems, observability refers to the possibility of deducing the latent variables in a system from observation of the manifest variables. In this way, these important concepts are extended far beyond the classical state space setting.

We view control as the design of a subsystem in an interconnected system, a subsystem that interacts with the plant through certain pre-specified variables, the control variables. For a linear time-invariant differential plant, it is possible to prove that a behavior is implementable by a linear time-invariant controller if and only if its behavior is wedged in between the hidden behavior and the realizable plant behavior.

The pre-occupation of systems and control with input/output systems does not do proper justice to the nature of physical systems: most physical systems are simply not a signal processors. Notwithstanding the importance of signal processors, the universal view of a system as an input/output device is simply a *faux pas*. And an unneccesary one at that: the behavioral approach offers a viable alternative.

## References

[1] J.W. Polderman, J.C. Willems, Introduction to Mathematical Systems Theory: A Behavioral Approach, Springer, Berlin, 1998.

[2] J.C. Willems, System theoretic models for the analysis of physical systems, Ricerche di Automatica 10 (1979) 71–106.

[3] J.C. Willems, From time series to linear system — Part I. Finite dimensional linear time invariant systems, Automatica 22 (1986) 561–580.

[4] J.C. Willems, From time series to linear system — Part II. Exact modelling, Automatica 22 (1986) 675–694.

[5] J.C. Willems, From time series to linear system — Part III. Approximate modelling, Automatica 23 (1987) 87–115.

[6] J.C. Willems, Models for dynamics, Dynamics Reported 2 (1989) 171–269.

[7] J.C. Willems, Paradigms and puzzles in the theory of dynamical systems, IEEE Trans. Automatic Contr. 36 (1991) 259–294.

[8] J.C. Willems, On interconnections, control, and feedback, IEEE Trans. Automatic Contr. 42 (19997) 326–339.

[9] H.K. Pillai, S. Shankar, A behavioral approach to control of distributed systems, SIAM J. Contr. Optimizat. 37 (1999) 388–408.

[10] J.C. Willems, Open dynamical systems and their control, in: Proceedings of the International Conference of Mathematicians, Berlin, Documenta Mathematica, Volume ICM 1998 — invited papers, 1998, pp. 697–706.

[11] P. Rocha, J.C. Willems, Controllability of 2-D systems, IEEE Trans. Automatic Contr. 36 (19991) 413–423.

[12] P. Rocha, J.C. Willems, Behavioral controllability of delay-differential Systems, SIAM J. Contr. Optimizat. 35 (1997) 254–264.

[13] U. Oberst, Multidimensional constant linear systems, Acta Applicandae Mathematicae 20 (1990) 1–175.

[14] H. Glüsing-Lüerssen, A behavioral approach to delay-differential systems, SIAM J. Contr. Optimizat. 35 (1997) 480–499.

[15] M. Fliess, S.T. Glad, An algebraic approach to linear and nonlinear control, in: H.L. Trentelman, J.C. Willems (Eds.), Essays on Control: Perspectives in the Theory and Its Applications, Birkhäuser, 1993, pp. 223–267.