THE BEHAVIORAL APPROACH TO SYSTEMS THEORY

Jan C. Willems *

Abstract

The behavioral approach provides a mathematical framework for modeling, analysis, and synthesis of dynamical systems. The main difference from the classical view is that it does not take an input/output partition as its starting point. Further, it incorporates latent variables in addition to the manifest variables which the model aims at. State variables appear as a special case of a latent variable model. This framework fits the practice of modeling by tearing and zooming, particularly useful for dealing with interconnected physical systems.

1 Introduction

The purpose of my lecture is to outline a mathematical language for modeling, analysis, and synthesis of systems. The framework that I 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 view which has dominated the development of the field of systems and control in the 20-th 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.

It is not the purpose to develop mathematical ideas for their own sake. To the contrary, I will downplay mathematical issues of a technical nature. The main aim is to demonstrate that the behavioral framework is a cogent systems-theoretic setting that properly deals with physical systems and that uses modeling as the essential motivation for choosing appropriate mathematical concepts. The behavioral framework aims at models of systems that interact with their environment. However, while systems that process information are very important in engineering practice and often call for an input/output setting, it is simply not logical to view (the interconnection of) physical systems in this way.

It is impossible to do justice to this theory in one lecture. I will therefore concentrate of a few main themes. The behavioral approach is discussed, including mathematical technicalities, in many articles (see [7], where additional references may be found), and in the textbook [3].

2 The behavior

The framework that I will use for discussing mathematical models views a model as follows. Assume that we have a phenomenon that we wish to model. Nature (that is, the

*ESAT, K.U. Leuven, B-3001 Leuven, Belgium, Jan.Willems@esat.kuleuven.ac.be.

reality that governs this phenomenon) can produce certain events (also called outcomes). The totality of possible events (*before* we have modeled the phenomenon) forms a set $\mathbb{U}$, called the *universum*. A *mathematical model* of the phenomenon restricts the outcomes that are declared possible to a subset $\mathscr{B}$ of $\mathbb{U}$; $\mathscr{B}$ is called the *behavior* of the model. I refer to $(\mathbb{U},\mathscr{B})$ (or to $\mathscr{B}$ by itself, since $\mathbb{U}$ usually follows from the context) as a mathematical model.

Examples:

1. *The port behavior of an electrical resistor.* The outcomes are pairs (V,I) with V the voltage (say, in volts) across the resistor and I the current (say, in amps) through the resistor. The universum is $\mathbb{R}^2$. After the resistor is modeled (say, by Ohm's law), the behavior is $\mathscr{B} = \{(V,I) \in \mathbb{R}^2 | V = RI\}$ with R the value of the resistance (say, in ohms).

2. Economists believe that in a *production process* there is a relation between the total production P, the labor L performed, and the capital K invested. The universum $\mathbb{U}$ equals $(\mathbb{R}_+)^3$ (taking into account that P,K,L are non-negative quantities). A possible model for this production process is $P = \alpha L^\beta K^\gamma$ with $\alpha,\beta,\gamma \in \mathbb{R}_+$ *technology coefficients* (assumed constant for a specific production process). Whence $\mathscr{B} = \{(P,L,K) \in (\mathbb{R}_+)^3 | P = \alpha L^\beta K^\gamma\}$.

3. The *ideal gas law* poses $PV = kNT$ as the relation between the pressure P, the volume V, the number N of moles, and the temperature T of an ideal gas, with k a universal physical constant. The universum $\mathbb{U}$ is $(\mathbb{R}_+)^4$, and the behavior $\mathscr{B} = \{(P,V,N,T) \in (\mathbb{R}_+)^4 | PV = kNT\}$.

In the study of dynamical systems we are, more specifically, interested in situations where the events are signals, trajectories, i.e. maps from a set of *independent variables* (time, or space, or time and space) to a set of *dependent variables* (the values taken on by the signals). In this case the universum is the collection of all maps from the set of independent variables to the set of dependent variables. It is convenient to distinguish these sets explicitly in the notation: $\mathbb{T}$ for the set of independent variables, and $\mathbb{W}$ for the set of dependent variables. $\mathbb{T}$ suggests 'time', but in distributed parameter systems $\mathbb{T}$ is often time and space. I have incorporated distributed systems because of their importance in engineering models. Whence a (dynamical) *system* is defined as a triple

$$\boxed{\Sigma = (\mathbb{W},\mathbb{T},\mathscr{B})}$$

with $\mathscr{B}$, *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 $\mathbb{W}$). The behavior is the central object in this definition. It formalizes which signals $w : \mathbb{T} \to \mathbb{W}$ are possible, according to the model: those in $\mathscr{B}$, and which are not: those not in $\mathscr{B}$.

Examples:

1. *Newton's second law* imposes a restriction that relates the position $\vec{q}$ of a point mass with mass m to the force $\vec{F}$ acting on it: $\vec{F} = m\frac{d^2}{dt^2}\vec{q}$. This is a dynamical system with $\mathbb{T} = \mathbb{R}, \mathbb{W} = \mathbb{R}^3 \times \mathbb{R}^3$ (typical elements of $\mathscr{B}$ are maps $(\vec{q},\vec{F}) : \mathbb{R} \to \mathbb{R}^3 \times \mathbb{R}^3$), and behavior $\mathscr{B}$ consisting of all maps $t \in \mathbb{R} \mapsto (\vec{q},\vec{F})(t) \in \mathbb{R}^3 \times \mathbb{R}^3$ that satisfy $\vec{F} = m\frac{d^2}{dt^2}\vec{q}$. I do not specify the precise sense of what it means that a function satisfies a differential equation (I will pay no attention to such secondary issues).

2. One-dimensional *diffusion* describes the evolution of the temperature $T(x,t)$ (with $x \in \mathbb{R}$ position and $t \in \mathbb{R}$ time) along a uniform bar and the heat $q(x,T)$ supplied to it. Their relation is given by the partial differential equation

$$\frac{\partial}{\partial t}T = \frac{\partial^2}{\partial x^2}T + q$$

where the constants are assumed to have been chosen appropriately. This defines a system with $\mathbb{T} = \mathbb{R}^2$, $\mathbb{W} = \mathbb{R}^2$, and $\mathscr{B}$ consisting of all maps $(T,q) : \mathbb{R}^2 \to \mathbb{R}^2$ that satisfy this partial differential equation.

3. *Maxwell's equations* provide 'the' example of a distributed system with many independent variables. They describe the possible realizations of the electric field $\vec{E} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$, the magnetic field $\vec{B} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$, the current density $\vec{j} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$, and the charge density $\rho : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$ in free space. Maxwell's equations are

$$\nabla \cdot \vec{E} = \frac{1}{\varepsilon_0}\rho, \qquad \nabla \times \vec{E} = -\frac{\partial}{\partial t}\vec{B},$$
$$\nabla \cdot \vec{B} = 0, \qquad c^2 \nabla \times \vec{B} = \frac{1}{\varepsilon_0}\vec{j} + \frac{\partial}{\partial t}\vec{E},$$

with ε_0 the dielectric constant of the medium and c^2 the speed of light in the medium. This defines the system $(\mathbb{R} \times \mathbb{R}^3, \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}, \mathscr{B})$, with $\mathscr{B}$ the set of all fields $(\vec{E}, \vec{B}, \vec{j}, \rho) : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}$ that satisfy Maxwell's (partial differential) equations.

4. *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 $\mathscr{B}$ the set of maps $w : \mathbb{R} \to \mathbb{R}^3$ that satisfy Kepler's laws: 1. the orbits 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; 2. the radius vector from the sun to the planet must sweep out equal areas in equal time, and 3. the ratio of the period of revolution around the ellipse to the major axis must be the same for all w's in $\mathscr{B}$.

These examples fit perfectly the notion of a dynamical system as a triple $\Sigma = (\mathbb{T}, \mathbb{W}, \mathscr{B})$ with $\mathscr{B} \subseteq \mathbb{W}^{\mathbb{T}}$. In examples 1 and 4, the set of independent variables $\mathbb{T}$ is time only, while in the second example, diffusion, and in the third, Maxwell's equations, $\mathbb{T}$ involves time and space. Note that in each of these examples, except the last, we are dealing with 'open' systems, that is, systems that interact with their environment (mathematically, systems in which appropriate initial conditions are insufficient to determine the solution uniquely). It has been customary to deal with such systems by viewing them as input/output systems, and by assuming that the input is imposed by the environment. Of course, the first two examples could be thought of as input/output systems. This requires some flexibility in the case of Newton's second law in order to avoid getting caught in a debate of causality in mechanics. In the case of diffusion, the heat supplied may be thought of as caused by an external heating mechanism that imposes q. But q may also be the consequence of radiation due to the temperature of the bar, making the assumption that it is q that causes the evolution of T untenable, since it is more like T that causes the radiation of heat. It is inappropriate to force Maxwell's equations (where there are clearly free variables in the system: the number of equations, 8, being strictly less than the number of variables, 10) into an input/output setting.

The input/output setting imposes an unnecessary - usually unphysical and often needlessly confusing - signal flow structure on systems in interaction with their environment. The input/output point of view has many virtues as a vehicle of studying physical systems, but as a starting point, it is simply inappropriate. First principles laws in physics always state that some events 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 from the outside by free inputs which together with an initial state specifies the other variables, the outputs. The behavioral framework treats a model for what it is: an exclusion law.

3 Latent variables

In the basic equations describing systems, very often other variables appear in addition to those whose behavior the model aims at describing. The origin of these auxiliary variables varies from case to case. They may be state variables (as in automata and input/state/output systems); they may be potentials (as in the well-known expressions for the solutions of Maxwell's equations); most frequently, they are interconnection variables (see section 5). It is important to incorporate these variables in our modeling language *ab initio*, and to distinguish clearly between the variables whose behavior the model aims at, and the auxiliary variables introduced in the modeling process. I call the former *manifest* variables, and the latter *latent* variables.

A *mathematical model with latent variables* is defined as a triple $(\mathbb{U}, \mathbb{L}, \mathscr{B}_{\text{full}})$ with $\mathbb{U}$ the universum of manifest variables, $\mathbb{L}$ the universum of latent variables, and $\mathscr{B}_{\text{full}} \subseteq \mathbb{U} \times \mathbb{L}$ the *full behavior*. It induces the *manifest model* $(\mathbb{U}, \mathscr{B})$, with $\mathscr{B} = \{w \in \mathbb{U} \mid \text{there exists } \ell \in \mathbb{L} \text{ such that } (w, \ell) \in \mathscr{B}_{\text{full}}\}$. A (dynamical) *system with latent variables* is defined completely analogously as

$$\boxed{\Sigma_{\text{full}} = (\mathbb{T}, \mathbb{W}, \mathbb{L}, \mathscr{B}_{\text{full}})}$$

with $\mathscr{B}_{\text{full}} \subseteq (\mathbb{W} \times \mathbb{L})^{\mathbb{T}}$. The notion of a system with latent variables is the natural end-point of a modeling process and hence a very natural starting point for the analysis and synthesis of systems. Latent variables enter also very forcefully in representation questions.

Examples:

1. In modeling the port behavior of a *1-port electrical circuit*, the manifest variables are the voltage V across and the current I into the circuit through the port. However, it is usually not possible to come up directly with a model (say in the form of a differential equation) that involves only the port variables (V, I). In order to model the port behavior, we usually need to look at the internal structure of the circuit, and introduce the currents through and the voltages across the internal branches as latent variables. Using Kirchhoff's laws and the constitutive equations of the elements in the branches, this readily leads to a latent variable model. Similar situations occur in other systems, for example mechanical systems, and more generally any type of interconnected system.

2. Assume that we want to model the relation between *the temperatures and the heat flows* radiated at the ends of a uniform bar of length 1. The bar is assumed to be isolated, except at the ends. Denote the manifest variables by T_0, T_1, q_0, q_1. In order to obtain a model, it is convenient to introduce the temperature distribution $T(x,t), 0 \leq x \leq 1$, in the bar as latent variables. The full behavior is then described by the partial differential equation

$$\frac{\partial}{\partial t} T = \frac{\partial^2}{\partial x^2} T$$

with the boundary conditions

$$T(0,t) = T_0(t), T(1,t) = T_1(t), \frac{\partial}{\partial x} T(0,t) = -q_0(t), \frac{\partial}{\partial x} T(1,t) = q_1(t).$$

This defines a latent variable system with $\mathbb{T} = \mathbb{R}, \mathbb{W} = \mathbb{R}^4, \mathbb{L} = \mathscr{C}^\infty([0,1], \mathbb{R})$, and $\mathscr{B}_{\text{full}}$ consisting of all maps $((T_0, T_1, q_0, q_1), T) : \mathbb{R} \to \mathbb{R}^4 \times \mathscr{C}^\infty([0,1], \mathbb{R}))$ such that the above equations are satisfied.

3. In a *state model* for a dynamical system, the input/output behavior is specified the state evolution law f and the read-out map h through

$$\dot{\mathrm{x}} = f(\mathrm{x}, \mathrm{u}), \quad \mathrm{y} = h(\mathrm{x}, \mathrm{u}).$$

This defines the latent variable system $(\mathbb{R}, \mathbb{U} \times \mathbb{Y}, \mathbb{X}, \mathscr{B}_{\text{full}})$ with $\mathscr{B}_{\text{full}}$ consisting of all trajectories $((u,y),x) : \mathbb{R} \to (\mathbb{U} \times \mathbb{Y}) \times \mathbb{X}$ that satisfy $\frac{d}{dt}x = f \circ (x,u), y = h \circ (x,u)$. The manifest behavior is the input/output behavior, that is all trajectories $(u,y) : \mathbb{R} \to \mathbb{U} \times \mathbb{Y}$ that are 'supported' (in the sense made apparent by the full behavior) by a trajectory $x : \mathbb{R} \to \mathbb{X}$.

Situations in which models use latent variables either for mathematical reasons or in order to express the behavioral constraints abound: internal voltages and currents in electrical circuits, momenta in mechanics, entropy and internal energy in thermodynamics, prices in economics, state variables, the wave function in quantum mechanics in order to explain observables, the basic probability space Ω, etc.

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 $\mathscr{B}$, i.e. a set of possible trajectories $w : \mathbb{T} \to \mathbb{W}$, that is the central object of study. However, as we have seen, in first principles modeling, also latent variables enter *ab initio*. But, the sets $\mathscr{B}$ or $\mathscr{B}_{\text{full}}$ of trajectories must be specified somehow, and it is here that differential equations (and difference equations in discrete-time systems) 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 by and large differential equations are certainly the most prevalent specification of behaviors encountered in applications.

For systems described by ODE's (*1-D* systems), with $\mathbb{T} = \mathbb{R}$, and in the case without latent variables, $\mathscr{B}$ consists of the solutions of a system of differential equations as

$$f_1 \circ (w, \frac{d}{dt}w, \ldots, \frac{d^N}{dt^N}w) = f_2 \circ (w, \frac{d}{dt}w, \ldots, \frac{d^N}{dt^N}w).$$

Call these *1-D differential systems*. In the case of systems with latent variables these differential equations involves both manifest and latent variables, yielding

$$f_1 \circ (w, \frac{d}{dt}w, \ldots, \frac{d^N}{dt^N}w, \ell, \frac{d}{dt}\ell, \ldots, \frac{d^N}{dt^N}\ell) = f_2 \circ (w, \frac{d}{dt}w, \ldots, \frac{d^N}{dt^N}w, \ell, \frac{d}{dt}\ell, \ldots, \frac{d^N}{dt^N}\ell),$$

as the equation relating the (vector of) manifest variables w to the (vector of) latent variables ℓ.

Of particular interest (in control, signal processing, circuit theory, econometrics, etc.) are systems with a signal space that is a finite-dimensional vector space and behavior described by linear constant coefficient differential (or difference) equations. Such systems occur not only when the dynamics are linear, but also after linearization around an equilibrium point, when studying the 'small signal behavior'. A *1-D linear time-invariant differential system* is a dynamical system $\Sigma = (\mathbb{R}, \mathbb{W}, \mathscr{B})$, with $\mathbb{W} = \mathbb{R}^{\mathtt{w}}$ a finite-dimensional (real) vector space, whose behavior consists of the solutions of a system of differential equations of the form

$$R_0 w + R_1 \frac{d}{dt}w + \cdots + R_n \frac{d^n}{dt^n}w = 0,$$

with $R_0, R_1, \ldots, R_n$ matrices of appropriate size that specify the system parameters, and $w = (w_1, w_2, \ldots, w_{\mathtt{w}})$ the vector of (real-valued) system variables. These systems call for polynomial matrix notation. It is convenient to denote the above system of differential equations as

$$\boxed{R(\tfrac{d}{dt})w = 0,}$$

with $R \in \mathbb{R}^{\bullet \times \mathtt{w}}[\xi]$ a real polynomial matrix with $\mathtt{w}$ columns. The behavior of this system is defined as

$$\mathscr{B} = \{w : \mathbb{R} \to \mathbb{R}^{\mathtt{w}} \mid R(\frac{d}{dt})w = 0\}.$$

The precise definition of what are solutions of $R(\frac{d}{dt})w = 0$ is an issue that I will slide over, but for the results that follow, it is convenient to consider solutions in $\mathscr{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{w}})$. Since $\mathscr{B}$ is the kernel of the differential operator $R(\frac{d}{dt})$, I often write $\mathscr{B} = \ker\big(R(\frac{d}{dt})\big)$, and call $R(\frac{d}{dt})w = 0$ a *kernel representation* of the associated linear time-invariant differential system. The set of differential systems or their behaviors is denoted by $\mathscr{L}^\bullet$, or by $\mathscr{L}^{\mathtt{w}}$ when the number of variables is $\mathtt{w}$.

Of course, the number of columns of the polynomial matrix R equals the dimension of $\mathbb{W}$. The number of rows of R, which represents the number of scalar differential equations, is arbitrary. In fact, when the row dimension of R is less than its column dimension, as is usually the case, then $R(\frac{d}{dt})w = 0$ is an under-determined system of differential equations which is typical for models in which the influence of the environment is taken into account.

In the case of linear time-invariant differential systems with latent variables, this leads to

$$\boxed{R(\tfrac{d}{dt})w = M(\tfrac{d}{dt})\ell,}$$

with R and M polynomial matrices of appropriate sizes. Define the *full* behavior of this system as

$$\mathscr{B}_{\text{full}} = \{(w, \ell) : \mathbb{R} \to \mathbb{R}^{\mathtt{w} \times \ell} \mid R(\frac{d}{dt})w = M(\frac{d}{dt})\ell\}.$$

Hence the *manifest* behavior of this system is

$$\mathscr{B} = \{w : \mathbb{R} \to \mathbb{R}^{\mathtt{w}} \mid \text{ there exists } \ell : \mathbb{R} \to \mathbb{R}^{\ell} \text{ such that } R(\frac{d}{dt})w = M(\frac{d}{dt})\ell\}.$$

$R(\frac{d}{dt})w = M(\frac{d}{dt})\ell$ is called a *latent variable* representation of the manifest behavior $\mathscr{B}$.

There is a very extensive theory about these linear differential systems. It is a natural first class of systems for a theory of dynamical systems. Besides being a natural outcome of modeling (perhaps after linearization), it incorporates high order differential equations, the ubiquitous first order state systems and transfer function models, implicit (descriptor) systems, etc., all as special cases. In fact, in a typical application, with a model of an interconnected system, many of the equations in $R(\frac{d}{dt})w = M(\frac{d}{dt})\ell$ are indeed algebraic equations. The study of these systems is intimately connected with the study of polynomial matrices, and may seem somewhat abstract and unfamiliar. Some of the central mathematical results obtained are:

1. *Mathematical structure:* there is a $1 \leftrightarrow 1$ correspondence between $\mathscr{L}^{\mathtt{w}}$ and the sub-modules of $\mathbb{R}^{\mathtt{w}}[\xi]$ viewed as a module over $\mathbb{R}[\xi]$. This $1 \leftrightarrow 1$ correspondence associates with each $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$ its set of *annihilators* $\mathscr{N}_{\mathscr{B}}$, i.e. the linear differential relations (sometimes called *consequences*) that annihilate $\mathscr{B}$. The set $\mathscr{N}_{\mathscr{B}} \subseteq \mathbb{R}^{\mathtt{w}}[\xi]$ consists of the polynomial vectors $n \in \mathbb{R}^{\mathtt{w}}[\xi]$ such that $n^\top(\frac{d}{dt})\mathscr{B} = 0$, and is easily seen to be a sub-module of $\mathbb{R}^{\mathtt{w}}[\xi]$. This $1 \leftrightarrow 1$ relationship brings many issues in the study of $\mathscr{L}^\bullet$ back to the structure of these sub-modules.

2. *Inclusion:* the behavior of $R_1(\frac{d}{dt})w = 0$ is contained in that of $R_2(\frac{d}{dt})w = 0$ iff $R_2 = FR_1$ with F a polynomial matrix of suitable dimensions.

3. *Minimality of kernel representations:* every $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$ has a kernel representation $R(\frac{d}{dt})w = 0$ with R a polynomial matrix of full row rank. Such kernel representations are called *minimal*, because of all kernel representations of a given $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$, the number of rows of R (i.e. the number of scalar differential equations defining $\mathscr{B}$) is a small as possible. A kernel representation is minimal iff the transposes of the rows of R form an independent set of generators of $\mathscr{N}_{\ker\left(R(\frac{d}{dt})\right)}$.

4. *Uniqueness of minimal kernel representations:* two minimal kernel representations $R_1(\frac{d}{dt})w = 0$ and $R_2(\frac{d}{dt})w = 0$ represent the same system iff $R_1 = UR_2$ with U a unimodular polynomial matrix.

5. *Elimination:* the question occurs whether the manifest behavior $\mathscr{B}$ of a latent variable representation belongs to $\mathscr{L}^{\mathtt{w}}$, i.e. whether it can also be described by a linear constant coefficient differential equation. This is called the *elimination problem*. The following *elimination theorem* holds: 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 $R(\frac{d}{dt})w = M(\frac{d}{dt})\ell$ has the kernel representation $R'(\frac{d}{dt})w = 0$.

6. *Input/output partition:* it is easy to prove that a *1-D* linear differential system admits an *input/output representation*. This means that for every $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$, there exists a permutation matrix Π and a partition $\Pi w = \mathrm{col}(u,y)$ such that for any $u^* \in \mathscr{C}^\infty(\mathbb{R},\mathbb{R}^{\mathtt{u}})$, there exist a $y \in \mathscr{C}^\infty(\mathbb{R},\mathbb{R}^{\mathtt{y}})$ such that $(u^*,y) \in \Pi\mathscr{B}$. Moreover, the y's that such $(u^*,y) \in \Pi\mathscr{B}$ form a linear finite dimensional variety, implying that such a y is uniquely determined by its derivatives at $t = 0$. Thus in linear differential systems, the variables can always be partitioned into two groups. The first group acts a free inputs, the second group as bound outputs: they are completely determined by the inputs and their initial conditions. The associated behavioral equations are

$$P(\frac{d}{dt})y = Q(\frac{d}{dt})u, \quad w = \Pi \begin{bmatrix} u \\ y \end{bmatrix},$$

with P square, $\det(P) \neq 0$, (one can even obtain $P^{-1}Q$ proper), and Π a suitably chosen permutation matrix that serves to partition the variables in inputs and outputs.

The relevance of the elimination problem in object-oriented modeling is as follows. A model obtained by tearing and zooming usually involves very many auxiliary (latent) variables and very many equations, among them many algebraic ones originating from the interconnection constraints. The elimination theorem tells us that (for *1-D* linear time-invariant differential systems) the latent variables may be completely eliminated and 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 will go up in the elimination process. There exist very effective, computer algebra based, algorithms for going from a latent variable representation to a kernel representation. Various generalizations of the elimination theorem and of elimination algorithms to other classes of systems (for example, time-varying or certain classes of nonlinear systems) have been obtained.

Now, let us consider briefly the generalization of some of the above concepts and results to constant coefficient *PDE*'s. Define a $\mathtt{n}$-*D* distributed linear shift-invariant differential system as a system $\Sigma = (\mathbb{R}^{\mathtt{n}},\mathbb{R}^{\mathtt{w}},\mathscr{B})$, whose behavior $\mathscr{B}$ consists of the $(\mathscr{C}^\infty(\mathbb{R}^{\mathtt{n}},\mathbb{R}^{\mathtt{d}}))$ solutions of a system of linear constant-coefficient partial differential equations

$$\boxed{R(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_{\mathtt{n}}})w = 0}$$

viewed as an equation in the w's, in the functions

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

Here, $R \in \mathbb{R}^{\bullet \times \mathtt{w}}[\xi_1, \ldots, \xi_{\mathtt{n}}]$ is a matrix of polynomials in $\mathbb{R}[\xi_1, \ldots, \xi_{\mathtt{n}}]$, polynomials with real coefficients in $\mathtt{n}$ indeterminates. In many applications, for example in Maxwell's equations, $\mathtt{n} = 4$ and the independent variables are (t, x, y, z), for time and space. Important properties of these systems are their *linearity* (meaning that $\mathscr{B}$ is a linear subspace of $(\mathbb{R}^{\mathtt{w}})^{\mathbb{R}^{\mathtt{n}}}$), and *shift-invariance* (meaning $\sigma^x \mathscr{B} = \mathscr{B}$ for all $x \in \mathbb{R}^{\mathtt{n}}$, where σ^x denotes the $x-$shift, defined by $(\sigma^x f)(x') = f(x' + x)$). *PDE* is called a *kernel representation* of the $\mathtt{n}$-*D* system $\Sigma = (\mathbb{R}^{\mathtt{n}}, \mathbb{R}^{\mathtt{w}}, \mathscr{B})$ with $\mathscr{B}$ the set of solutions of $R(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})w = 0$. These systems, or their behavior, are denoted by $\mathscr{L}_{\mathtt{n}}^{\bullet}$, and by $\mathscr{L}_{\mathtt{n}}^{\mathtt{w}}$ when there are $\mathtt{n}$ dependent variables.

For distributed differential systems with latent variables, this leads to PDE's of the form

$$\boxed{R(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})w = M(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})\ell,}$$

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

$$\mathscr{B}_{\text{full}} = \{(w, \ell) \mid \text{ the above } PDE \text{ in } (w, \ell) \text{ holds}\}$$

and the *manifest behavior* as

$$\mathscr{B} = \{w \mid \text{ there exists } \ell \text{ such that } (w, \ell) \in \mathscr{B}_{\text{full}}\}$$

The *PDE* with latent variables is called a *latent variable* representation of $\mathscr{B}$.

Some of the results for *1-D* differential systems have recently been generalized to $\mathtt{n}$-*D* differential systems [4]. The study of these systems is intimately connected with the study of polynomial matrices in $\mathtt{n}$ variables. The follwoing are some of the central mathematical results that have been obtained.

1. *Mathematical structure:* there is a bijective correspondence between $\mathscr{L}_{\mathtt{n}}^{\mathtt{w}}$ and the submodules of $\mathbb{R}^{\mathtt{w}}[\xi_1, \ldots, \xi_{\mathtt{n}}]$ viewed as a module over $\mathbb{R}^{\mathtt{w}}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$. This correspondence associates with each $\mathscr{B} \in \mathscr{L}_{\mathtt{n}}^{\mathtt{w}}$ the $n \in \mathbb{R}^{\mathtt{w}}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$ such that $n^T(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})\mathscr{B} = 0$.

2. *Inclusion:* the behavior of the system $R_1(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})w = 0$ is contained in that of $R_2(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})w = 0$ iff $R_2 = FR_1$ with F a polynomial matrix of suitable dimensions.

3. *Elimination:* the question occurs again whether the manifest behavior $\mathscr{B}$ of an $\mathtt{n}$-*D* latent variable representation belongs to $\mathscr{L}_{\mathtt{n}}^{\mathtt{w}}$, i.e. whether it can also be described by a linear constant coefficient partial differential equation. This is the case indeed. The following *elimination theorem* holds. For any real polynomial matrices (R, M) with $\text{rowdim}(R) = \text{rowdim}(M)$, there exists a real polynomial matrix R' such that the manifest behavior of $R(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})w = M(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})\ell$ has kernel representation $R'(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})w = 0$.

The elimination theorem for *1-D* systems is rather easy to prove. It follows readily from the Smith canonical form for real polynomial matrices in one indeterminate. However, the elimination theorem for $\mathtt{n}$-*D* system is much more difficult to prove. It is a consequence of the *Fundamental Principle*, a deep result in the theory of linear *PDE's*. This states the following. Consider the system of *PDE's*

$$P(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})x = y$$

with $P \in \mathbb{R}^{g_1 \times g_2}[\xi_1, \ldots, \xi_{\mathtt{n}}]$ and $y \in \mathscr{C}^\infty(\mathbb{R}^{\mathtt{n}}, \mathbb{R}^{g_1})$ given, and $x \in \mathscr{C}^\infty(\mathbb{R}^{\mathtt{n}}, \mathbb{R}^{g_2})$ the unknown. The *Fundamental Principle* states that this system of *PDE's* is solvable iff the *left syzygies* of P annihilate y, i.e. iff $n \in \mathbb{R}^{g_1}[\xi_1, \ldots, \xi_{\mathtt{n}}]$, and $n^\top P = 0$ imply $n^\top(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})y = 0$.

There exist very effective, Gröbner basis-based, algorithms for elimination, i.e. for computing R' from (R,M).

As an illustration of the elimination theorem for *PDE's*, consider the elimination of $\vec{B}$ and ρ from Maxwell's equations. The elimination theorem tells us that the $(\vec{E},\vec{j})$ that are declared possible by Maxwell's equations themselves satisfy a linear *PDE*. The following equations describe the possible realizations of the fields $\vec{E}$ and $\vec{j}$:

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

5 Tearing and zooming

Systems, especially engineering systems, usually consist of interconnections of subsystems. This feature is crucial in both modeling and design. The procedure of modeling by *tearing and zooming* is an excellent illustration of the appropriateness of the behavioral approach as the supporting mathematical language. Assume throughout finiteness, i.e. that we interconnect a finite number of modules (subsystems), each with a finite number of terminals, etc. See figure 1.

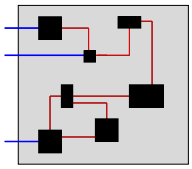

Figure 1: Interconnected system

I view an interconnected system as a collection of *modules* with *terminals*, interconnected through an *interconnection architecture*. The building blocks, called *modules*, of an interconnected system are systems with *terminals*. Each of these terminals carries variables from a universum, and the (dynamical) laws that govern the module are expressed by a behavior that relates the variables at its terminals. Finally, the terminals of the modules are assumed to be interconnected, expressed by an *interconnection architecture*. The interconnection architecture also imposes relations between the variables on these terminals.

After interconnection, the architecture leaves some terminals available for interaction with the environment of the overall system. The behavior of the interconnected system consists of the signals that satisfy both the behavioral laws of the modules and the interconnection constraints. In specifying the behavior of an interconnected system, we can often consider the variables on the interconnected terminals as latent variables, and those on the terminals that are left for interaction with the environment as manifest variables. In this sense, we view the interconnection variables as *internal* variables, and the exposed variables *external* variables. It is important to note immediately the hierarchical nature of this procedure. The modules thus become subsystems. The paradigmatic example to keep in mind is an electrical circuit. The modules are resistors, capacitors, inductors, transformers, etc. The terminals are the wires attached to the modules and are electrical terminals, each carrying a voltage (the potential) and a current. The interconnection architecture states how the wires are connected.

Modeling interconnected via the procedure of *tearing and zooming* provides the prime example of the usefulness of behaviors and the inadequacy of input/output thinking. Even if the system, after interconnection, allows for a natural input/output representation, it is

unlikely that this will be the case of the subsystem and of the interconnection architecture. This is illustrated by means of an example.

Example: Consider two tanks filled with a fluid, both equipped with two tubes through which the fluid can flow in or out (see figure 2). Assume that the pressures (p_{11}, p_{12}) and the flows (f_{11}, f_{12}) at the end of these tubes of the first tank tank are governed by a differential equation of the form

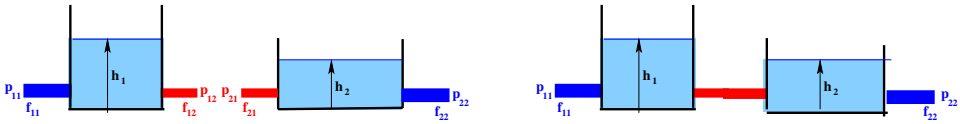

Figure 2: Two tanks

$$\frac{d}{dt}h_1 = F_1(h_1, p_{11}, p_{12}), \;\; f_{11} = H_{11}(h_1, p_{11}), \;\; f_{12} = H_{12}(h_1, p_{12}),$$

where h_1 denotes the height of the fluid in the first tank. Similarly for the second tank:

$$\frac{d}{dt}h_2 = F_2(h_2, p_{21}, p_{22}), \;\; f_{21} = H_{21}(h_2, p_{21}), \;\; f_{22} = H_{22}(h_2, p_{22})$$

It is quite reasonable, by all accounts, to consider in the first system the pressures p_{11}, p_{12} as inputs, the flows f_{11}, f_{12} as outputs, and h_1 as state, and for the second system the pressures p_{21}, p_{22} as inputs, the flows f_{21}, f_{22} as outputs, and h_2 as state. Now assume that we interconnect the tube 12 to 21. This yields the interconnection laws:

$$p_{12} = p_{21}, f_{12} + f_{21} = 0.$$

Note that this comes down to equating two inputs, and equating two outputs. Precisely the opposite that what is supposed to happen in the output-to-input identification that signal flow modeling and SIMULINK© wants us to do! A similar situation holds in mechanics: interconnection equates two positions (often both outputs), and puts the sum of two forces (often both inputs) equal to zero, and in many many other circumstances.

If the field of systems and control wants to take modeling seriously, is should retrace the *faux pas* of taking input/output thinking as the basic framework, and cast models in the language of behaviors. It is only when considering the more detailed signal flow graph structure of a system that input/output thinking becomes useful. Signal flow graphs are useful building blocks for interpreting information processing systems, but not at all for physical systems, and certainly not for interconnection, which need a more flexible framework.

6 Controllability and observability

An important property in the analysis and synthesis of 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 $\dot{\mathrm{x}} = f(\mathrm{x}, \mathrm{u})$ is said to be controllable if for all states $a, b \in \mathbb{X}$, there exist a time $T \geq 0$ and an input $u : [0, T] \to \mathbb{U}$ and such that the solution to $\frac{d}{dt}x = f \circ (x, u)$ and $x(0) = a$ yields $x(T) = b$.

One of the elementary results of system theory states that the finite-dimensional linear system $\dot{x} = Ax + Bu$ is controllable iff 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 this classical notion of controllability as formulated above is that it refers to a particular representation of a system, notably to a state 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 manifest system variables directly.

Let $\Sigma = (\mathbb{T}, \mathbb{W}, \mathscr{B})$ be a dynamical system with $\mathbb{T} = \mathbb{R}$ or $\mathbb{Z}$, and assume that is time-invariant, that is $\sigma^t \mathscr{B} = \mathscr{B}$ for all $t \in \mathbb{T}$, where σ^t denotes the t-shift ($(\sigma^t f)(t') := f(t' + t)$); Σ is said to be *controllable* if for all $w_1, w_2 \in \mathscr{B}$ there exists $T \in \mathbb{T}$, $T \geq 0$ and $w \in \mathscr{B}$ 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. This is illustrated in figure 3.

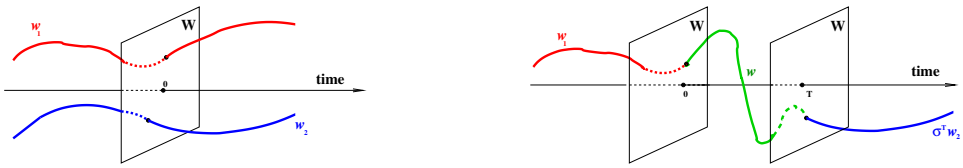

Figure 3: Controllability

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. Let $\Sigma = (\mathbb{R}, \mathbb{R}^{\mathtt{w}}, \mathscr{B})$ be a linear time-invariant differential system. The following are equivalent:

1. $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$ is controllable.

2. The polynomial matrix R in a kernel representation $R(\frac{d}{dt})w = 0$ of $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$ satisfies $\operatorname{rank}\big(R(\lambda)\big) = \operatorname{rank}(R)$ for all $\lambda \in \mathbb{C}$, i.e. $\operatorname{rank}\big(R(\lambda)\big)$ is the same for all $\lambda \in \mathbb{C}$.

3. The behavior $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$ is the image of a linear constant-coefficient differential operator, that is, there exists a polynomial matrix $M \in \mathbb{R}^{\mathtt{w} \times \bullet}[\xi]$ such that $\mathscr{B} = \{w \mid w = M(\frac{d}{dt})\ell$ for some $\ell\}$.

4. The compact support elements of $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$ are dense in $\mathscr{B}$.

5. The $\mathbb{R}[\xi]$-module $\mathbb{R}^{\mathtt{w}}[\xi]/\mathscr{N}_{\mathscr{B}}$, where $\mathscr{N}_{\mathscr{B}}$ denotes the annihilators of $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$, is torsion free.

6. The submodule $\mathscr{N}_{\mathscr{B}}$ is *'slim'*, meaning that it contains no submodules of the same dimension. The *dimension* of a submodule of $\mathbb{R}^{\mathtt{w}}[\xi]$ equals the minimal number of independent generators.

There exist various algorithms for verifying controllability of a system starting from the coefficients of the polynomial matrix R in a kernel (or (R, M) in a latent variable) representation of Σ. The basic idea is to compute *syzygies* associated with R. These are standard algorithms in computer algebra packages. I will however not enter into these

Figure 4: Observability

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

$$\boxed{w = M(\tfrac{d}{dt})\ell.}$$

This is called an *image* representation of the system with behavior

$$\mathscr{B} = \{w \mid \text{ there exists } \ell \text{ such that } w = M(\frac{d}{dt})\ell\}.$$

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 are precisely the controllable ones for which this is possible.

The controllability question has been pursued for many other classes of systems. In particular (more difficult to prove) generalizations have been derived for differential-delay systems [5, 2], for nonlinear systems, and, as will be discussed soon, for *PDE's*. Image representation have received much attention for nonlinear differential-algebraic systems, where they are referred to as *flat* systems [1]. Flatness implies controllability.

The notion of observability is always introduced hand in hand with controllability. In the context of the input/state/output system $\dot{\mathrm{x}} = f(\mathrm{x},\mathrm{u}), \mathrm{y} = h(\mathrm{x},\mathrm{u})$, it refers to the possibility of deducing, using the laws of the system, the state trajectory from observation of the input and output trajectories. The definition that is used in the behavioral context is more generally applicable in that the variables that are observed and the variables that need to be deduced are kept arbitrary.

The question that is asked in observability is the following: Can the trajectory w_1 be deduced from the trajectory w_2? (see figure 4). Let $\Sigma = (\mathbb{T}, \mathbb{W}, \mathscr{B})$ 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 Σ if $(w_1, w_2') \in \mathscr{B}$ and $(w_1, w_2'') \in \mathscr{B}$ imply $w_2' = w_2''$. Observability thus refers to the possibility of deducing w_1 from observation of w_2 and the laws of the system ($\mathscr{B}$ is assumed to be known).

The further theory of observability runs parallel to that of controllability. I 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 (i.e. a set of consequences) of the following form, that puts observability into evidence: $w_1 = R_2'(\frac{d}{dt})w_2$. This condition is again readily turned into a standard problem in computer algebra.

Many of the results for controllability and observability have recently been generalized to distributed systems [4]. I mention them briefly. The system $\mathscr{B} \in \mathscr{L}_{\mathtt{n}}^{\mathtt{w}}$ is said to be *controllable* if for all $w_1, w_2 \in \mathscr{B}$ and for all open non-overlapping subsets $O_1, O_2 \subseteq \mathbb{R}^{\mathtt{n}}$,

there exists $w \in \mathscr{B}$ such that $w|_{O_1} = w_1|_{O_1}$ and $w|_{O_2} = w_2|_{O_2}$, i.e. if its solutions are 'patch-able'. This is illustrated in figure 5.

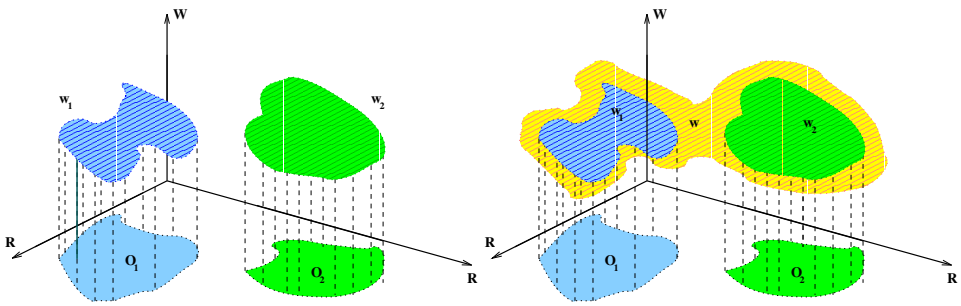

Figure 5: Controllability of PDE's

Note that it follows from the elimination theorem for $\mathscr{L}_{\mathtt{n}}^{\bullet}$ that the manifest behavior of a system in image representation, i.e. a latent variable system of the special form

$$\boxed{w = M(\tfrac{\partial}{\partial x_1}, \ldots, \tfrac{\partial}{\partial x_{\mathtt{n}}})\ell}$$

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. It turns out that it are precisely the controllable systems that admit an image representation. Indeed in [4], it is shown that the following statements are equivalent for systems described by constant coefficient linear *PDE's*:

1. $\mathscr{B} \in \mathscr{L}_{\mathtt{n}}^{\mathtt{w}}$ defines a controllable system.

2. $\mathscr{B} \in \mathscr{L}_{\mathtt{n}}^{\mathtt{w}}$ admits an image representation.

3. The trajectories of compact support are dense in $\mathscr{B} \in \mathscr{L}_{\mathtt{n}}^{\mathtt{w}}$.

4. The $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$-module $\mathbb{R}^{\mathtt{w}}[\xi_1, \ldots, \xi_{\mathtt{n}}]/\mathscr{N}_{\mathscr{B}}$, where $\mathscr{N}_{\mathscr{B}}$ denotes the annihilators of $\mathscr{B} \in \mathscr{L}_{\mathtt{n}}^{\mathtt{w}}$, is torsion free.

Note that an image representation corresponds to what in mathematical physics is called a *potential function*, with ℓ the potential and $M(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_{\mathtt{n}}})$ 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 (patch-ability of trajectories in the behavior).

It can be shown that Maxwell's equations define a controllable distributed differential system. Indeed, in the case of Maxwell's equations, a well-known image representation is given by

$$\vec{E} = -\frac{\partial}{\partial t}\vec{A} - \nabla\phi, \qquad \vec{j} = \varepsilon_0 \frac{\partial^2}{\partial t^2}\vec{A} - \varepsilon_0 c^2 \nabla^2 \vec{A},$$
$$\vec{B} = \nabla \times \vec{A}, \qquad \rho = \frac{\varepsilon_0}{c^2}\frac{\partial^2}{\partial t^2}\phi - \varepsilon_0 \nabla^2 \phi,$$

where $\phi : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$ is a scalar, and $\vec{A} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$ a vector potential. The existence of such an image representation is equivalent to controllability, i.e. to patch-ability of solutions to Maxwell's equations.

For *1-D* systems it is easy to show that every controllable linear time-invariant differential behavior $\mathscr{B}$ admits an observable image representation (meaning that in $\mathscr{B}_{\text{full}}$ the latent variables are observable from the manifest ones). This, however, does not hold for $\mathtt{n}$-*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 $M(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_{\mathtt{n}}})\ell = 0$ has solutions $\ell \neq 0$. This means that however one represents a controllable behavior $\mathscr{B}$ of a *PDE* as $w = M(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_{\mathtt{n}}})\ell$, there does not exist an $M' \in \mathbb{R}^{\mathtt{w}\times\bullet}[\xi_1,\ldots,\xi_{\mathtt{n}}]$ such that $w = M(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_{\mathtt{n}}})\ell$ implies $\ell = M'(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_{\mathtt{n}}})w$. The latent variables will then not be recoverable from the manifest ones. For example, the image representation of the behavior defined by Maxwell's equations in terms of the scalar potential ϕ and the vector potential $\vec{A}$. In fact, Maxwell's equations are an example of a controllable system that does not allow an observable image representation.

7 State representations

As shown in the previous sections, there are many ways to represent a system. Further, as we have seen, systems are often modeled with latent variables and this obviously induces a multitude of descriptions of a given manifest behavior. One very important system representation is the state space representation. It is this class of system representations that induced a revolution in systems and control theory in the early sixties. I will now explain, in the context of systems with independent variable only time, how the state can be viewed in the context of latent variable systems in which the latent variable has the additional property that it separates the past from the future.

Let $\Sigma = (\mathbb{T},\mathbb{W},\mathbb{X},\mathscr{B}_{\text{full}})$ be a latent variable system with $\mathbb{T} \subseteq \mathbb{R}$. Then this systems is said to be a *state system* if $\mathscr{B}_{\text{full}}$ has the following property: if $(w_1,x_1) \in \mathscr{B}_{\text{full}}$, $(w_2,x_2) \in \mathscr{B}_{\text{full}}$, and $x_1(t_0) = x_2(t_0)$, then concatenation of (w_1,x_1) and (w_2,x_2) at t_0 must also belong to $\mathscr{B}_{\text{full}}$. Concatenation is defined as

$$((w_1,x_1) \wedge_{t_0} (w_2,x_2))(t) = \begin{cases} (w_1,x_1)(t) & \text{for} \quad t < t_0 \\ (w_2,x_2)(t) & \text{for} \quad t \leq t_0 \end{cases}$$

This property must hold for all such (w_1,x_1), $(w_2,x_2) \in \mathscr{B}$, and all $t_0 \in \mathbb{T}$ (see figure 6).

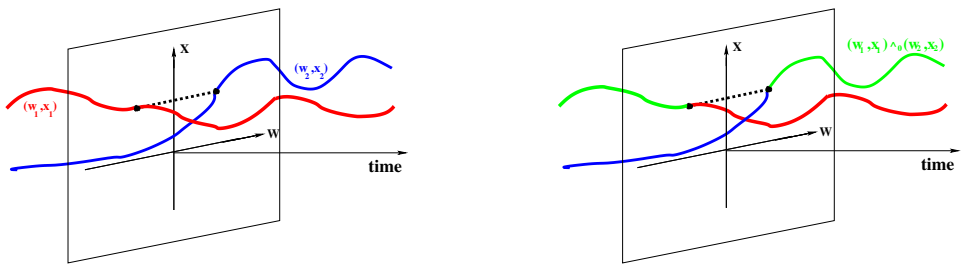

Figure 6: State property

It is possible to prove, under some mild regularity conditions, that a differential system is a state space system iff $\mathscr{B}_{\text{full}}$ can be described by a differential equation that is *first order* in x and *zero-th order* in w, i.e. a differential equation of the form

$$f_1 \circ (w,x,\frac{d}{dt}x) = f_2 \circ (w,x,\frac{d}{dt}x)$$

In particular, for linear time-invariant differential systems, this yields the representation

$$\boxed{Kw + Lx + M\tfrac{d}{dt}x = 0}$$

with K, L, M matrices of suitable dimensions.

It can be shown that for every system described by $R(\frac{d}{dt})w$ (and hence, by the elimination theorem, for every system described by linear constant-coefficient differential equations with latent variables) it is possible to define (new, or additional) latent variables $x = X(\frac{d}{dt})w$ such that the relation restricting the w's and the x's is of the above state space form, and such that its manifest behavior is $\ker\big(R(\frac{d}{dt})\big)$. One such *state map* $X \in \mathbb{R}^{\bullet\times\mathtt{w}}[\xi]$ can be constructed as follows: $X = \mathrm{col}(\sigma R, \sigma^2 R, \ldots)$, where the *cut-and-shift* operator σ associates with the polynomial (matrix) $p(\xi) = p_0 + p_1\xi + p_2\xi^2 + \cdots$, the polynomial (matrix) $(\sigma p)(\xi) := p_1 + p_2\xi + p_3\xi^2 + \cdots$. A similar very effective state construction holds for systems in image or latent variable representation.

Moreover, if the dimension of x (equivalently, the number of rows of X) is as small as possible, under the constraint that the relation between x and w is of the above first order form and that the manifest behavior is $\ker\big(R(\frac{d}{dt})\big)$, then the above first order state equation can be written as

$$\boxed{\tfrac{d}{dt}x = Ax + Bw_1, \quad w_2 = Cx + Dw_1}$$

with (w_1, w_2) a component-wise partition of w, i.e. $\mathrm{col}(w_1, w_2) = \Pi w$ with Π a suitable permutation matrix. This shows how the popular input/state/output representation emerges as a special representation that involves (i) a suitable, component-wise, partition of the manifest variables, and (ii) a suitable choice of the latent variables to represent the behavior in state form. Hence such an input/state/output representation displays both the partition of its manifest variables in its free (input) components and its bound (output) components, and also its memory (as parameterized by the state variables). The usefulness of such representations for the analysis and synthesis is uncontested. However, as a *starting point* for practical and theoretical work, it is not suitable. It requires a great deal of manipulations and computation to bring a first principles model into this form. This is already apparent for linear constant-coefficient differential systems, and this is all the more so for more general classes of systems, for example nonlinear ones.

8 Conclusions

In this paper, I have covered some highlights of the behavioral approach to systems and control. A mathematical model is defined 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 modeling process). The central object in behavioral systems theory is a system with latent variables.

Various problems occur in this framework. In particular, 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 patch-ability 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 their classical state space setting.

Because of space limitations, I have been unable to discuss control in a behavioral setting. Control is viewed 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 two behaviors: the hidden behavior and the realizable plant behavior [6].

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.

Acknowledgments

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 Modelling, by the KUL Concerted Research Action (GOA) MEFISTO–666, and by several grants en projects from IWT-Flanders and the Flemish Fund for Scientific Research.

References

[1] M. Fliess and S.T. Glad, An algebraic approach to linear and nonlinear control, pages 223-267 of *Essays on Control: Perspectives in the Theory and Its Applications*, edited by H.L. Trentelman and J.C. Willems, Birkhäuser, 1993.

[2] H. Glüsing-Lüerssen, A behavioral approach to delay-differential systems, *SIAM Journal on Control and Optimization*, volume 35, pages 480-499, 1997.

[3] J.W. Polderman and J.C. Willems, *Introduction to Mathematical Systems Theory: A Behavioral Approach,* Springer-Verlag, 1998.

[4] H.K. Pillai and S. Shankar, A behavioral approach to control of distributed systems, *SIAM Journal on Control and Optimization*, volume 37, pages 388-408, 1999.

[5] P. Rocha and J.C. Willems Behavioral controllability of delay-differential Systems, *SIAM Journal on Control and Optimization*, volume 35, pages 254-264, 1997.

[6] J.C.Willems, On interconnections, control, and feedback, *IEEE Transactions on Automatic Control*, volume 42, pages 326-339, 1997.

[7] `www.esat.kuleuven.ac.be/~jwillems`