The Behavioral Approach to Modeling and Control of Dynamical Systems

Jan C. Willems[1]
Mathematics Institute
University of Groningen
The Netherlands

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 the input/output partition as its starting point. In this setting, control is viewed as interconnection.

Keywords

Behaviors, Tearing and zooming, Controllability, Control as interconnection

Introduction

The purpose of this paper is to outline the basics of a mathematical language for the modeling, analysis, and the synthesis of systems. The framework that we 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 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, we will downplay mathematical issues of a technical nature. 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 uses modeling as the essential motivation for choosing appropriate mathematical concepts.

It is impossible to do justice to all these aspects in the span of one article. We will therefore concentrate of a few main themes. The behavioral approach is discussed, including the mathematical technicalities, in the recent textbook (Polderman and Willems, 1998), where additional references may be found. We also mention the article (Pillai and Shankar, 1999) where some of these results are generalized to partial differential equations.

The Behavior

The framework that we 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 reality that governs this phenomenon) can produce certain events (we will also call them outcomes). The totality of these possible events (*before* we have modeled the phenomenon) forms a set $\mathbb{W}$, called the *universum.* A *mathematical model* of this phenomenon restricts the outcomes that are declared possible to a subset $\mathfrak{B}$ of $\mathbb{W}$; $\mathfrak{B}$ is called the *behavior* of the model. We refer to $(\mathbb{W}, \mathfrak{B})$ (or to $\mathfrak{B}$ by itself, since $\mathbb{W}$ is usually obvious 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, by Ohm's law, the behavior is $\mathfrak{B} = \{(V, I) \in \mathbb{R}^2 | V = RI\}$ with R the value of the resistance (say, in ohms).

2. 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 physical constant. The universum $\mathbb{W}$ is $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{N} \times \mathbb{R}_+$, and the behavior $\mathfrak{B} = \{(P, V, N, T) \in \mathbb{W} | PV = kNT\}$.

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 for a mathematical model: $\mathbb{T}$ 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 - we have incorporated distributed systems because of their importance in chemical engineering models. Whence we define a *system* as a triple

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

with $\mathfrak{B}$, *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 $\mathfrak{B}$, and which are not: those not in $\mathfrak{B}$.

[1] willems@math.rug.nl

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 $\mathfrak{B}$ are maps $(\vec{q}, \vec{F}) : \mathbb{R} \to \mathbb{R}^3 \times \mathbb{R}^3$), and behavior $\mathfrak{B}$ consisting of all maps $t \in \mathbb{R} \mapsto (\vec{q}, \vec{F})(t) \in \mathbb{R}^3 \times \mathbb{R}^3$ that satisfy $\vec{F} = m\frac{d^2}{dt^2}\vec{q}$. We do not specify the precise sense of what it means that a function satisfies a differential equation (we will pay almost 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 $\mathfrak{B}$ 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 $\rho : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$. Maxwell's equations are

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

   with ε_0 the dielectric constant of the medium and c^2 the speed of light in the medium. This defines the system $(\mathbb{R} \times \mathbb{R}^3, \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}, \mathfrak{B})$, with $\mathfrak{B}$ 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.

These examples fit perfectly our notion of a dynamical system as a triple $\Sigma = (\mathbb{T}, \mathbb{W}, \mathfrak{B})$ with $\mathfrak{B} \subseteq \mathbb{W}^{\mathbb{T}}$. In example 1 , 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, 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, our first two examples could be thought of as input/output systems. 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 - and unphysical - signal flow structure on our view of systems in interaction with our 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.

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 (we will discuss this later). It is important to incorporate these variables in our basic 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. We call the former *manifest* variables, and the latter *latent* variables.

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

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

with $\mathfrak{B}_{\text{full}} \subseteq (\mathbb{W} \times \mathbb{L})^{\mathbb{T}}$. 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. We shall see that latent variables enter also very forcefully in representation questions.

Examples

1. In modeling the port behavior of an *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 (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. We wish to model the relation between $q_0,\ T_0,\ q_1,\ T_1$, the heat flows and temperatures at both ends. 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
$$T(0,t) = T_0(t), \qquad T(1,t) = T_1(t),$$
$$\frac{\partial}{\partial x}T(0,t) = -q_0(t), \qquad \frac{\partial}{\partial x}T(1,t) = q_1(t).$$
This defines a latent variable system with $\mathbb{T} = \mathbb{R}, \mathbb{W} = \mathbb{R}^4, \mathbb{L} = \mathfrak{C}^\infty([0,1],\mathbb{R})$, and $\mathfrak{B}_{\text{full}}$ consisting of all maps $((T_0,T_1,q_0,q_1),T) : \mathbb{R} \to \mathbb{R}^4 \times \mathfrak{C}^\infty([0,1],\mathbb{R}))$ such that the above equations are satisfied.
3. In a state model for a dynamical system, the input/output behavior is specified through a system of differential equations as
$$\frac{d}{dt}x = f(x,u), \quad y = h(x,u).$$
This defines a latent variable system $(\mathbb{R}, \mathbb{U} \times \mathbb{Y}, \mathbb{X}, \mathfrak{B}_{\text{full}})$ with $\mathfrak{B}_{\text{full}}$ all trajectories $((u,y),x) : \mathbb{R} \to (\mathbb{U} \times \mathbb{Y}) \times \mathbb{X}$ that satisfy these equations. 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, chemical potentials, 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 in probability, etc.

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 $\mathfrak{B}$, 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 $\mathfrak{B}$ or $\mathfrak{B}_{\text{full}}$ 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, hybrid systems), but we do not consider these very relevant refinements in the present paper.

For systems described by ODE's (*1-D* systems), with $\mathbb{T} = \mathbb{R}$, and in the case without latent variables, $\mathfrak{B}$ consists of the solutions of a system of differential equations as
$$f_1(w, \frac{d}{dt}w, \ldots, \frac{d^N}{dt^N}w) = f_2(w, \frac{d}{dt}w, \ldots, \frac{d^N}{dt^N}w).$$
We call these *1-D differential systems*. In the case of systems with latent variables these differential equations involve both manifest and latent variables, yielding
$$f_1(w, \frac{d}{dt}w, \ldots, \frac{d^N}{dt^N}w, \ell, \frac{d}{dt}\ell, \ldots, \frac{d^N}{dt^N}\ell) = f_2(w, \frac{d}{dt}w, \ldots, \frac{d^N}{dt^N}w, \ell, \frac{d}{dt}\ell, \ldots, \frac{d^N}{dt^N}\ell),$$
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 $\Sigma = (\mathbb{R}, \mathbb{W}, \mathfrak{B})$, 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

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

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

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

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

$$\{(w, \ell) : \mathbb{R} \to \mathbb{R}^{\mathtt{w}+\ell} \mid R(\frac{d}{dt})w = M(\frac{d}{dt})\ell\}.$$

Hence the *manifest* behavior of this system is

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

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

There is a very extensive theory about these linear differential systems. It is a natural starting point for a theory of dynamical systems. Besides being the 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., as special cases. The study of these systems is intimately connected with the study of polynomial matrices, and may seem somewhat abstract, but this is only because of unfamiliarity. See (Polderman and Willems, 1998) for details.

An important issue that occurs is *elimination:* the question whether the manifest behavior $\mathfrak{B}$ of a latent variable representation belongs to $\mathfrak{L}^{\mathtt{w}}$, i.e., whether it can also be described by a linear constant coefficient differential equation. 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 $R(\frac{d}{dt})w = M(\frac{d}{dt})\ell$ has the kernel representation $R'(\frac{d}{dt})w = 0$.

The relevance of the elimination problem in object-oriented modeling is as follows. As we will see, 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. We should also mention that there exist very effective, computer algebra based, algorithms for going from a latent variable representation to a kernel representation. The generalization of the elimination theorem and of elimination algorithms to other classes of systems (for example, time-varying or certain classes of nonlinear systems) is a matter of ongoing research. Particularly interesting is the generalization of some of the above concepts and results to systems described by constant coefficient linear *PDE*'s. Define a $\mathtt{n}$-*D* distributed linear shift-invariant differential system as a system $\Sigma = (\mathbb{R}^{\mathtt{n}}, \mathbb{R}^{\mathtt{w}}, \mathfrak{B})$, whose behavior $\mathfrak{B}$ consists of the $(\mathfrak{C}^\infty(\mathbb{R}^{\mathtt{n}}, \mathbb{R}^{\mathtt{d}}))$ solutions of a system of linear constant-coefficient partial differential equations

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

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.

For distributed differential systems with latent vari-

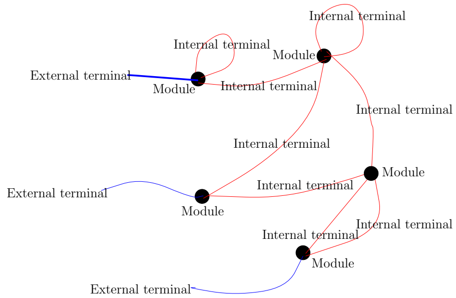

Figure 1: Interconnected system.

ables, this leads to equations 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 $\mathbb{R}[\xi_1, \ldots, \xi_{\mathtt{n}}]$.

It is easy to prove that a *1-D* linear differential system admits an *input/output representation*. This means that for every $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$, there exists a permutation matrix Π and a partition $\Pi w = \mathrm{col}(u, y)$ such that for any $u^* \in \mathfrak{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{u}})$, there exist a $y \in \mathfrak{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{y}})$ such that $(u^*, y) \in \Pi\mathfrak{B}$. Moreover, the y's that such $(u^*, y) \in \Pi\mathfrak{B}$ 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 act a free inputs, the second group a bound outputs: they are completely determined by the inputs and their initial conditions.

Tearing and Zooming

Systems, especially engineering systems, usually consist of interconnections of subsystems. This feature is crucial in both modeling and design. The aim of this section is to formalize interconnections and to analyze the model structures that emerge from it. The procedure of modeling by *tearing and zooming* is an excellent illustration of the appropriateness of the behavioral approach as the supporting mathematical language. We 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.

We 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 the various terminals. Finally, the terminals of the modules are assumed to be interconnected, expressed by an *interconnection architecture*. The interconnection architecture 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 module behavior laws and the interconnection constraints. In specifying the behavior of an interconnected system, we 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. We will occasionally call the interconnected variables *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. We now formalize all this, assuming that we are treating continuous time dynamical systems (hence, with time set $\mathbb{T} = \mathbb{R}$). Of course, for process engineering, generalization to distributed systems and to 'distributed' terminals, as in interconnection along surfaces, is mandatory.

A *terminal* is specified by its *type*. Giving the type of a terminal identifies the kind of a physical terminal that we are dealing with. The type of terminal implies a universum of *terminal variables*. These variables are physical quantities that characterize the possible 'signal states' on the terminal, it specifies how the module interacts with the environment through this terminal.

A *module* is specified by its *type*, and its *behavior*. Giving the type of a module identifies the kind of a physical system that we are dealing with. Giving a *behavior specification* of a module implies giving a *representation* and the values of the associated *parameters* a representation. Combined these specify the behavior of the variables on the terminals of the module. The type of a module implies an ordered set of terminals. Since each of the terminals comes equipped with a universum of terminal variables, we thus obtain an ordered set of variables associated with that module. The module behavior then specifies what time trajectories are possible for these variables. Thus a module defines a dynamical system $(\mathbb{R}, \mathbb{W}, \mathfrak{B})$ with $\mathbb{W}$ the Cartesian product over the terminals of the universa of the terminal variables. However, there are very many ways to specify a behavior (for example, as the solution set of a differential equation, as the image of a differential operator, through a latent variable model, through a transfer function, etc.). The behavioral representation picks out one of these. These representations will then contain unspecified parameters (for example, the coefficients of the differential equation, or the polynomials in a transfer function). Giving the parameter values specifies their numerical values, and completes the specification of the behavior of the signals that are possible on the terminals of a module.

Formally, a system Σ of a given type with T terminals yields $\mathbb{W} = \mathbb{W}_1 \times \mathbb{W}_2 \times \cdots \times \mathbb{W}_T$, with $\mathbb{W}_k$ the universum associated with the k-th terminal. The behavioral specification yields the behavior $\mathfrak{B} \subseteq \mathbb{W}^{\mathbb{R}}$. If $(w_1, w_2, \ldots, w_T) \in \mathfrak{B}$, then we think of $w_k \in (\mathbb{W}_k)^{\mathbb{R}}$ as a signal that can be realized on the k-th terminal.

An interconnected system is composed of modules, its building blocks. They serve as subsystems of the overall system. Each module specifies an ordered set of terminals. By listing the modules, and the associated terminals, we obtain the Cartesian product of all the terminals in the interconnected system. The manner in which these terminals, and hence the associated modules, are interconnected is specified by the *interconnection architecture*. This consists of a set of disjunct pairs of terminals, and it is assumed that each such pair consists of terminals of adapted type. Typical 'adapted' type means that the are the same physical nature (both electrical, or both 1-D mechanical, both thermal, etc. But, when the terminal serves for information processing (inputs to actuators, output of sensors) it could also mean that one variable must be an input to the module to which it is connected (say, the input of an actuator), and the other must be an output to the module to which it is connected (say the output of a sensor). Note that the interconnection architecture involves only the terminals of the modules and their type, but not the behavior. Also, the union of the terminals over the pairs that are part of the interconnection architecture will in general be a strict subset of the union of the terminals of all the modules. We call the terminals that are not involved in the interconnection architecture the *external (or exposed) terminals*. It is along these terminals that the interconnected system can interact with its environment. The terminals that enter in the interconnection architecture are called *internal terminal*. It is along these terminals that the modules are interconnected.

Pairing of terminals by the interconnection architecture implies an *interconnection law*. Some examples of interconnection laws are $V_1 = V_2, I_1 + I_2 = 0$ for electrical terminals, $Q_1 + Q_2 = 0, T_1 = T_2$ for thermal terminals, $p_1 = p_2, f_1 + f_2 = 0$ for fluidic terminals, etc.

The physical examples of interconnection laws all involve equating of 'across' variables and putting the sum of 'through' variables to zero. This is in contrast to the input-output identification for information processing terminals. The latter is actually the only interconnection that is used in flow diagram based modeling, as implemented, for example, in SIMULINK. It is indeed very much based on the input/output thinking that has permeated systems theory and control throughout the past century. Unfortunately, this is of limited interest

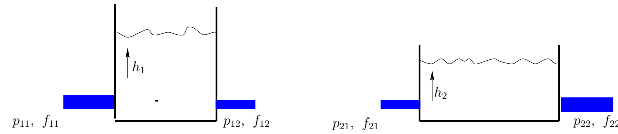

Figure 2: Tanks.

when it comes to modeling interconnected physical systems. As such the ideas developed in the *bond-graph* literature and the modeling packages that use this philosophy are bound to be much more useful in the long run. Interconnection of physical systems involves across and through variables, efforts and flows, extensive and intensive quantities, and not in first instance flow diagrams. These considerations are the main motivation for the development of the behavioral approach.

We now formalize the interconnected system. The most effective way to proceed is to specify it as a latent variable system, with as manifest variables the variables associated with the external terminals, and as latent variables the internal variables associated with the terminals that are paired by the interconnection architecture. This latent variable system is specified as follows. Its universum of manifest variables equals $\mathbb{W} = \mathbb{W}_{e_1} \times \cdots \times \mathbb{W}_{e_{|E|}}$, where $E = \{e_1, \ldots, e_{|E|}\}$ is the set of external terminal. Its universum of latent variables equals $\mathbb{L} = \mathbb{W}_{i_1} \times \cdots \times \mathbb{W}_{i_{|I|}}$, where $I = \{i_1, \ldots, i_{|I|}\}$ is the set of internal terminals. Its full behavior behavior consists of the behavior as specified by each of the modules, combined by the interconnection laws obtained by the interconnection architecture. The behavior of each of the modules involves a combination of internal and external variables that are associated with the module. The interconnection law of a pair in the interconnection architecture involves the internal variables associated with these terminals.

Modeling interconnected via the above method of *tearing and zooming* provides 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, it is unlikely that this will be the case of the subsystem and of the interconnection architecture. We illustrate this 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

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

where h_1 denotes the height of the fluid in the first tank.

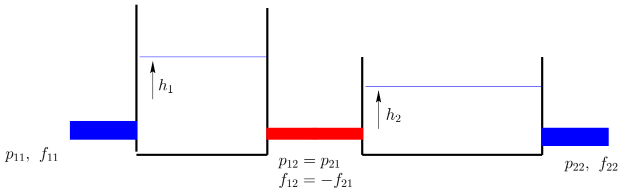

Figure 3: Connected tanks.

Similarly for the second tank:

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

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

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

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 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.

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 physical systems need a more flexible framework.

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 $\frac{d}{dt}x = f(x, u)$ is said to be controllable if for all states a, b, there exists an input$u$ and a time $T \geq 0$ such that the solution to $\frac{d}{dt}x = f(x, u)$ and $x(0) = a$ yields $x(T) = b$. One of the

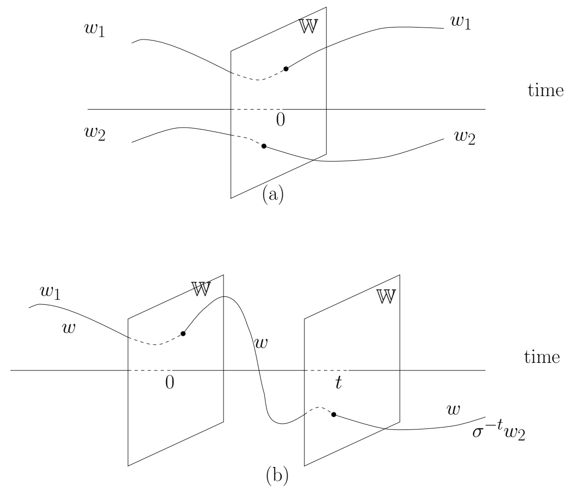

Figure 4: Controllability.

elementary results of system theory states that the finite-dimensional linear system $\frac{d}{dt}x = Ax + Bu$ 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 to 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 manifest system variables directly.

Let $\Sigma = (\mathbb{T}, \mathbb{W}, \mathfrak{B})$ be a dynamical system with $\mathbb{T} = \mathbb{R}$ or $\mathbb{Z}$, and assume that is time-invariant, that is $\sigma^t\mathfrak{B} = \mathfrak{B}$ for all $t \in \mathbb{T}$, where σ^t denotes the t-shift (defined by $(\sigma^t f)(t') := f(t' + t)$); Σ is said to be *controllable* if for all $w_1, w_2 \in \mathfrak{B}$ there exists $T \in \mathbb{T}$, $T \geq 0$ and $w \in \mathfrak{B}$ 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 4.

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

1. $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ is controllable.
2. The polynomial matrix R in a kernel representa-

tion $R(\frac{d}{dt})w = 0$ of $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ satisfies $\text{rank}(R(\lambda)) = \text{rank}(R)$ for all $\lambda \in \mathbb{C}$.

3. The behavior $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ 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 $\mathfrak{B} = \{w \mid w = M(\frac{d}{dt})\ell \text{ for some } \ell\}$.

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 Σ.

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

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

We call this an *image* representation of the system with behavior

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

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 (Rocha and Willems, 1997; Glüsing-Lüerssen, 1997), for nonlinear 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 (Fliess and Glad, 1993). Flatness implies controllability, but the exact relation remains to be studied.

The notion of observability is always introduced hand in hand with controllability. In the context of the input/state/output system $\frac{d}{dt}x = f(x,u), y = h(x,u)$, 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.

In observability, we ask the question: Can the trajectory w_1 be deduced from the trajectory w_2? (See figure 5). Let $\Sigma = (\mathbb{T}, \mathbb{W}, \mathfrak{B})$ 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 $(w_1, w_2') \in \mathfrak{B}$ and $(w_1, w_2'') \in \mathfrak{B}$ imply $w_2' = w_2''$. Observability thus refers to the possibility of deducing the w_1 from observation of w_2 and from the laws of the system ($\mathfrak{B}$ is assumed to be known).

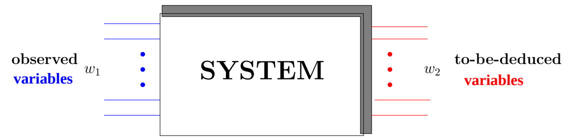

Figure 5: Observability.

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 (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 (Pillai and Shankar, 1999). We mention them briefly. The system $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}_{\mathtt{n}}$ is said to be *controllable* if for all $w_1, w_2 \in \mathfrak{B}$ and for all open non-overlapping subsets $O_1, O_2 \subseteq \mathbb{R}^{\mathtt{n}}$, there exists $w \in \mathfrak{B}$ 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'.

Note that it follows from the elimination theorem for $\mathfrak{L}^{\bullet}_{\mathtt{n}}$ 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 (Pillai and Shankar, 1999).

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, there exists a well-known image representation using the scalar and vector potential.

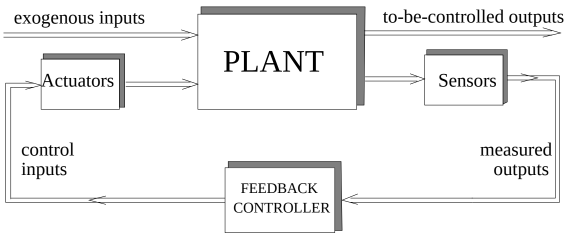

Figure 6: Intelligent control.

Control

The behavioral point of view has received broad acceptance as an approach for modeling dynamical systems. It is generally agreed upon that when modeling a dynamic component it makes no sense to prejudice oneself (as one would be forced to do in a transfer function setting) as to which variables should be viewed as inputs and which variables should be viewed as outputs. We have argued this point extensively throughout the previous sections of this paper. This is not to say, by any means, that there are no situations where the input/output structure is natural. Quite to the contrary. In fact, whenever logic devices are involved, in information processing, the input/output structure is often a must.

The behavioral approach has, until now, met with much less acceptance in the context of control. We can offer a number of explanations for this fact. Firstly, there is something very natural in viewing control variables as inputs and measured variables as outputs. Control then becomes decision making on the basis of observations. When subsequently a controller is regarded as a feedback processor, one ends up with the feeling that the input/output structure is in fact an essential feature of control. Secondly, since, as mentioned in a previous section, it is possible to prove that every linear time-invariant system admits a component-wise input/output representation, one gets the impression that the input/output framework can be adopted without second thoughts, that nothing is lost by taking it as a universal starting point.

Present-day control theory centers around the signal flow graph shown in figure 6. The plant has four terminals (supporting variables which will typically be vector-valued). There are two input terminals, one for the control, one for the other exogenous variables (disturbances, set-points, reference signals, tracking inputs, etc.) and there are two output terminals, one for the measurements, and one for the to-be-controlled variables. By using feed-through terms in the plant equations this configuration accommodates, by incorporating these variables in the measurements, for the possible availability to the controller of set-point settings, reference signals, or disturbance measurements for feed-forward control, and, by incorporating the control input in the to-be-controlled outputs, for penalizing excessive control action. The control inputs are generated by means of actuators and the measurements are made available through sensors. Usually, the dynamics of the actuators and of the sensors are considered to be part of the plant. We call this structure *intelligent control*. In intelligent control, the controller is thought of as a micro-processor type device which is driven by the sensor outputs and which produces the actuator inputs through an algorithm involving a combination of feedback, identification, and adaptation. Also, often loops expressing model uncertainty are incorporated in the above as well. Of course, many variations, refinements, and special cases of this structure are of interest, but the basic idea is that of supervisor reacting in an intelligent way to observed events and measured signals.

The paradigm embodied in figure 6 has been universally prevalent ever since the beginning of the subject, from our interpretation of the Watt regulator, Black's feedback amplifier, and Wiener's cybernetics, to the ideas underlying modern adaptive and robust control. It is indeed a deep and very appealing paradigm, which will undoubtedly gain in relevance and impact as logic devices become ever more prevalent, reliable, and inexpensive. This paradigm has a number of features which are important for considerations which will follow. Some of these are:

1. There is an asymmetry between the plant and the controller: it remains apparent what part of the system is the plant and what part is the controller. This asymmetry disappears to some extent in the closed loop.
2. The intelligent control paradigm tells us to be wary of errors in the measurements. Thus it is considered as being ill-advised to differentiate measurements, presumably, because this will lead to noise amplification.
3. The plant and the controller are dynamical systems which can be interconnected at any moment in time. If for one reason or another the feedback controller temporarily fails to receive an input signal, then the control input can be set to a default value, and later on the controller can resume its action.

We will now analyze an example of a down-to-earth controller, a very wide-spread automatic control mechanism, namely the traditional device which ensures the automatic closing of doors. There is nothing peculiar about this example. Devices based on similar principles are used for the suspension of automobiles and the points which we make through this example could also be made just as well through many temperature or pressure control devices. A typical automatic door-closing mechanism is schematically shown in figure 7.

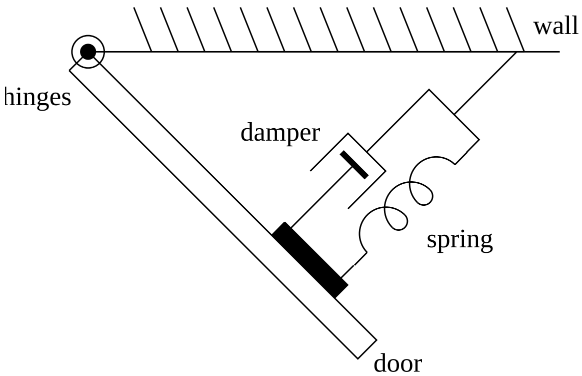

Figure 7: Door-closing mechanism.

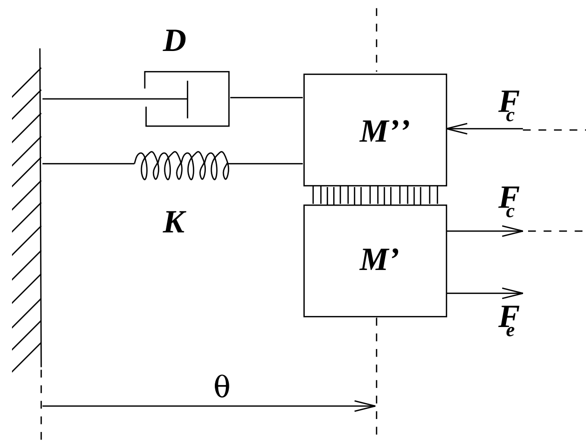

Figure 8: Spring/damper representation.

A door-closing mechanism usually consists of a spring to ensure the closing of the door and a damper in order to make sure that it closes gently. In addition, these mechanisms often have considerable weight so that their mass cannot be neglected as compared to the mass of the door itself. The automatic door-closing mechanism can be modeled as a mass/spring/damper combination. In good approximation, the situation can be analyzed effectively as the mechanical system shown in figure 8. We model the door as a mass M', on which, neglecting friction in the hinges, two forces act. The first force, F_c, is the force exerted by the door-closing device, while the second force, F_e, is the exogenous force exerted for example by a person pushing the door in order to open it. The equation of motion for the door becomes

$$M'\frac{d^2}{dt^2}\theta = F_c + F_e,$$

where θ denotes the opening angle of the door. The automatic door-closing mechanism, modeled as a mass/spring/damper combination, yields

$$M''\frac{d^2}{dt^2}\theta + D\frac{d}{dt}\theta + K\theta = -F_c.$$

Here, M'' denotes the mass of the door-closing mechanism, D its damping coefficient, and K its spring constant. Combining these equations yields

$$(M' + M'')\frac{d^2}{dt^2}\theta + D\frac{d}{dt}\theta + K\theta = F_e.$$

In order to ensure proper functioning of the door-closing device, the designer can to some extent choose M'', D and K (all of which must, for physical reasons, be positive). The desired response requirements are: small overshoot (to avoid banging of the door), fast settling time, and a reasonably high steady state gain (to avoid having to exert excessive force when opening the door). This is an example of an elementary control design exercise. A good design will be achieved by choosing a light mechanism (M'' small), with a reasonably strong spring (K large), but not too strong so as to avoid having to use excessive force in order to open the door, and with the value of D chosen so as to achieve slightly less than critical damping (ensuring gentle closing of the door).

It is completely natural to view in this example the door as the plant and the door-closing mechanism as the controller. Then, if we insist on interpreting this plant/controller combination in terms of control system configurations as figure 6, we obtain.

$$\text{Plant:}\quad M'\frac{d^2}{dt^2}\theta = u + v;\ y = \theta;\ z = \theta$$

with u the control input ($u = F_c$), v the exogenous input ($v = F_e$), y the measured output, and z the to-be-controlled output.

$$\text{Controller:}\quad u = -M''\frac{d^2}{dt^2}y - D\frac{d}{dt}y - Ky.$$

This yields the controlled system, described by:

$$\text{Controlled system:}\quad (M' + M'')\frac{d^2}{dt^2}z + D\frac{d}{dt}z + Kz = v.$$

Observe that in the control law, the measurement y should be considered as the input, and the control u should be considered as the output. This suggests that we are using what would be called a *PDD-controller* (a proportional and twice differentiating controller), a singular controller which would be thought of as causing high noise amplification. Of course, no such noise amplification occurs in reality. Further, the plant is second order, the controller is second order, and the closed loop system is also second order (unequal to the sum of the order of the plant and the controller). Hence, in order to connect the controller to the plant, we will have to 'prepare' the initial states of the controller and the plant. Indeed, in attaching the door-closing mechanism to the door, we will make sure that at the moment of attachment the initial values of θ and $\frac{d}{dt}\theta$ are zero *both* for the door and the door-closing mechanism.

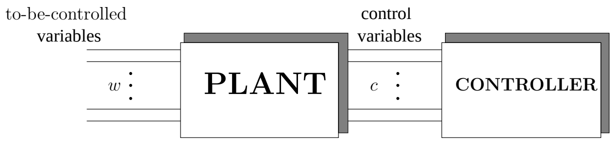

Figure 9: Control as interconnection.

We now come to our most important point concerning this example. Let us analyze the signal flow graph. In the plant, it is natural to view the forces F_c and F_e as inputs and θ as output. This input/output choice is logical both from the physical and from the cybernetic, control theoretic point of view. In the controller, on the other hand, the physical and the cybernetic points of view clash. From the cybernetic, control theoretic point of view, it is logical to regard the opening angle θ as input and the control force F_c as output. From the physical point of view, however, it is logical to regard (just as in the plant) the force F_c as input and θ as output. It is evident that as an interconnection of two mechanical systems, the door and the door-closing mechanism play completely symmetric roles. However, the cybernetic, control theoretic point of view obliges us to treat the situation as asymmetric, making the force the cause in one mechanical subsystem, and the effect in another.

In our opinion, this simple but realistic example permits us to draw the following conclusions. Notwithstanding all its merits, the intelligent control paradigm of figure 6 gives an unnecessarily restrictive view of control theory. In many practical control problems, the signal-flow-graph interpretation of figure 6 is untenable. The solution which we propose to this dilemma is the following. We will keep the distinction between plant and controller with the understanding that this distinction is justified only from an *evolutionary* point of view, in the sense that it becomes evident only *after* we comprehend the genesis of the controlled system, after we understand the way in which the closed loop system has come into existence as a purposeful system. However, we will abandon the intelligent control signal flow graph as a paradigm for control. We will abandon the distinction between control inputs and measured outputs. Instead, we will view *interconnection of a controller to a plant* as the central paradigm in control theory. However, we by no means claim that the intelligent control paradigm is without merits. To the contrary, it is extremely useful and important. Claiming that the input/output framework is *not always* the suitable framework to approach a problem does not mean that one claims that it is *never* the suitable framework. However, a good universal framework for control should be able to deal both with interconnection, with designing subsystems, and with intelligent control. The behavioral framework does, the intelligent control framework does not.

In order to illustrate the nature of control that we would like to transmit in this presentation, consider the system configuration depicted in figure 9. In the top part of the figure, there are two systems, shown as black-boxes with terminals. It is through their terminals that systems interact with their environment. The black-box imposes relations on the variables that 'live' on its terminals. These relations are formalized by the behavior of the system in the black-box. The system to the left in figure 9 is called the *plant*, the one to the right the *controller*. The terminals of the plant consist of *to-be-controlled variables* w, and *control variables* c. The controller has only terminals with the control variables c. In the bottom part of the figure, the control terminals of the plant and of the controller are connected. Before interconnection, the variables w and c of the plant have to satisfy the laws imposed by the plant behavior. But, after interconnection, the variables c also have to satisfy the laws imposed by the controller. Thus, after interconnection, the restrictions imposed on the variables c by the controller will be transmitted to the variables w. Choosing the black-box to the right so that the variables w have a desirable behavior in the interconnected black-box is, in our view, the basic problem of control. This point of view is discussed with examples in (Polderman and Willems, 1998).

This leads to the following mathematical formulation. The *plant* and the *controller* are both dynamical systems $\Sigma_{plant} = (\mathbb{T}, \mathbb{W} \times \mathbb{V}, \mathfrak{B}_{plant})$ and $\Sigma_{controller} = (\mathbb{T}, \mathbb{V}, \mathfrak{B}_{controller})$ where $\mathbb{W}$ denotes the signal space of the to-be-controlled variables, $\mathbb{V}$ denotes the signal space of control variables, and both systems are assumed to have the same time axis $\mathbb{T}$. The interconnection of Σ_{plant} and $\Sigma_{controller}$ leads to the system $\Sigma_{full} = (\mathbb{T}, \mathbb{W}, \mathbb{V}, \mathfrak{B}_{\text{full}})$ which is a system with latent variables ($\mathbb{V}$) and full behavior defined by

$$\mathfrak{B}_{\text{full}} = \{(w,c) : \mathbb{T} \to \mathbb{W} \times \mathbb{V} \mid (w,c) \in \mathfrak{B}_{plant} \text{ and } c \in \mathfrak{B}_{controller}\}$$

The manifest system obtained by Σ_{full} is the *controlled system* and is hence defined as $\Sigma_{controlled} = (\mathbb{T}, \mathbb{W}, \mathfrak{B}_{controlled})$ with

$$\mathfrak{B}_{controlled} = \{w : \mathbb{T} \to \mathbb{W} \mid \text{ there exists } c : \mathbb{T} \to \mathbb{V} \text{ such that} \in \mathfrak{B}_{plant} \text{ and } c \in \mathfrak{B}_{controller}\}$$

The notion of a controller put forward by the above view considers interconnection as the basic idea of control. The special controllers that consist of sensor-outputs to actuator-inputs signal processors emerge as (very important) special cases. We think of these as controllers as *feedback*, or *intelligent*, controllers. However, our view of control as the design of suitable subsystems greatly enhances the applicability of control, since it views control as integrated subsystem design.

A question that arises in this context is the following. Assume that Σ_{plant} is given. What systems $\Sigma_{controlled}$ can be obtained by suitably choosing $\Sigma_{controller}$? This question can be answered very explicitly, at least for linear time-invariant differential systems. Assume that the plant is given by $\Sigma_{plant} = (\mathbb{R}, \mathbb{R}^{\mathtt{w}} \times \mathbb{R}^{\mathtt{c}}, \mathfrak{B}_{plant}) \in \mathfrak{L}^{\mathtt{w}+\mathtt{c}}$ with $\mathfrak{B}_{plant}$ described by a system of linear constant differential equations $R(\frac{d}{dt})w = R(\frac{d}{dt})c$. Let $\Sigma_{controller} = (\mathbb{R}, \mathbb{R}_c, \mathfrak{B}_{controller}) \in \mathfrak{L}^{\mathtt{c}}$, and assume that $\mathfrak{B}_{controller}$ is similarly described by a system of linear constant coefficient differential equations $C(\frac{d}{dt})c = 0$. Then, by elimination theorem $\Sigma_{controlled} = (\mathbb{R}, \mathbb{R}^{\mathtt{w}}, \mathfrak{B}_{controlled})$ has also a behavior that is described by a system of linear constant coefficient differential equations. It turns out that the behaviors $\mathfrak{B}_{controlled}$ that can be obtained this way can be characterized very nicely.

Define therefore two behaviors, $\mathcal{P}$ and $\mathcal{N}$; $\mathcal{P}$ is called the *realizable (plant) behavior* and $\mathcal{N}$ the *hidden behavior*. They are defined as follows: $\mathcal{P}$ is the manifest behavior of the system, i.e.,

$$\boxed{\begin{array}{l}\mathcal{P} = \{w : \mathbb{R} \to \mathbb{R}^{\mathtt{w}} \mid \text{ there exists } c : \mathbb{R} \to \mathbb{R}^{\mathtt{c}} \\ \text{ such that } (w, c) \in \mathfrak{B}_{\text{full}}\},\end{array}}$$

and $\mathcal{N}$ is defined as

$$\boxed{\mathcal{N} = \{w : \mathbb{R} \to \mathbb{R}^{\mathtt{w}} \mid (w, 0) \in \mathfrak{B}_{plant}\}.}$$

Hence $\mathcal{N}$ is the behavior of the plant variables that are compatible with the control variables equal to zero. We say that $\mathfrak{B}_{controller}$ *implements* $\mathfrak{B}_{controlled}$ if there exists a controller such that the controlled behavior after interconnecting the controller with behavior $\mathfrak{B}_{controller}$ to the plant, yields $\mathfrak{B}_{controlled}$ as the controlled behavior.

The controller implementability problem asks what behaviors $\mathfrak{B}_{controlled}$ can be obtained this way. For linear time-invariant systems it is possible to prove that $\mathfrak{B}_{controlled}$ is implementable if and only if

$$\boxed{\mathcal{N} \subseteq \mathfrak{B}_{controlled} \subseteq \mathcal{P}.}$$

This result shows that the behaviors that are implementable by means of a controller are precisely those that are wedged in between the hidden behavior $\mathcal{N}$ and the realizable plant $\mathcal{P}$. The problem of control can therefore be reduced to finding such a behavior. Of course, the issue of how to construct $\Sigma_{controller}$ (for example, as a signal processor from the sensor outputs to the actuator inputs) must be addressed as well, but this can be done. In (Willems and Trentelman, 2001) this approach is used for the design of $\mathcal{H}_\infty$-controllers, and we discuss several results on the implementability by feedback controllers as well.

We believe that the point of view of control that emerges from this, as designing a subsystem (with feedback control as a special case) greatly enhances the scope and applicability of control as a discipline. In this setting, control comes down to sub-system design.

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 modeling process). The central object in behavioral systems theory is a system with latent variables.

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.

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

Fliess, M. and S. Glad, An algebraic approach to linear and nonlinear control, In Trentelman, H. and J. Willems, editors, *Essays on Control: Perspectives in the Theory and Its Applications*, pages 223–267. Birkhaäuser (1993).

Glüsing-Lüerssen, H., A behavioral approach to delay-differential systems, *SIAM Journal on Control and Optimization*, Volume 35, 480-499 (1997).

Pillai, H. and S. Shankar, A behavioral approach to control of distributed systems, *SIAM Journal on Control and Optimization*, Volume 37, 388-408 (1999).

Polderman, J. and J. Willems, *Introduction to Mathematical Systems Theory: A Behavioral Approach*. Springer-Verlag (1998).

Rocha, P. and J. Willems, Behavioral controllability of delay-differential Systems, *SIAM Journal on Control and Optimization*, Volume 35, 254-264 (1997).

Willems, J. and H. Trentelman, "Synthesis of dissipative systems using quadratic differential forms," *submitted to the IEEE Transactions on Automatic Control* (2001).