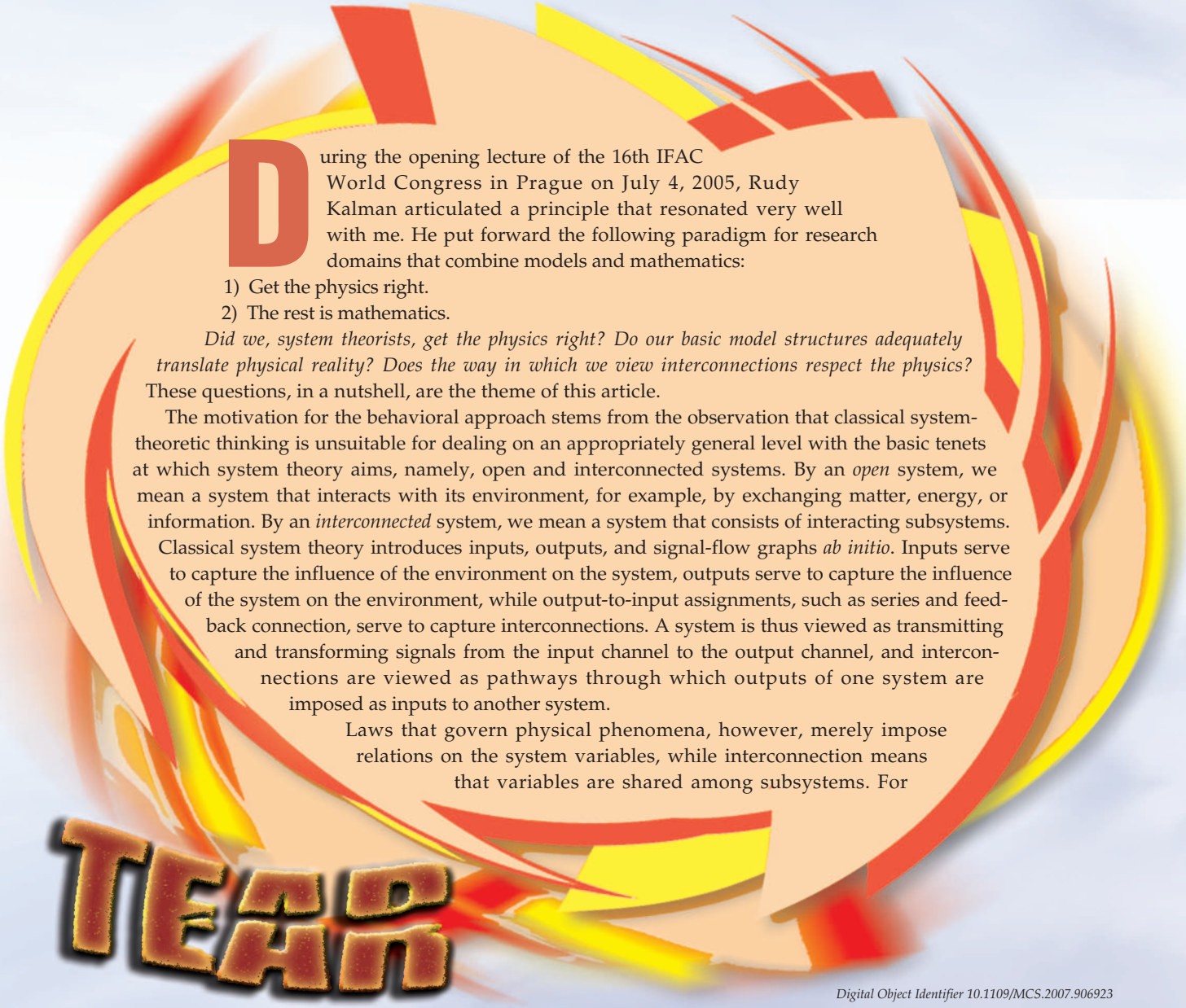
The Behavioral Approach to Open and Interconnected Systems

JAN C. WILLEMS

MODELING BY TEARING, ZOOMING, AND LINKING



During the opening lecture of the 16th IFAC World Congress in Prague on July 4, 2005, Rudy Kalman articulated a principle that resonated very well with me. He put forward the following paradigm for research domains that combine models and mathematics:

1) Get the physics right.
2) The rest is mathematics.

Did we, system theorists, get the physics right? Do our basic model structures adequately translate physical reality? Does the way in which we view interconnections respect the physics? These questions, in a nutshell, are the theme of this article.

The motivation for the behavioral approach stems from the observation that classical system-theoretic thinking is unsuitable for dealing on an appropriately general level with the basic tenets at which system theory aims, namely, open and interconnected systems. By an *open* system, we mean a system that interacts with its environment, for example, by exchanging matter, energy, or information. By an *interconnected* system, we mean a system that consists of interacting subsystems. Classical system theory introduces inputs, outputs, and signal-flow graphs *ab initio*. Inputs serve to capture the influence of the environment on the system, outputs serve to capture the influence of the system on the environment, while output-to-input assignments, such as series and feedback connection, serve to capture interconnections. A system is thus viewed as transmitting and transforming signals from the input channel to the output channel, and interconnections are viewed as pathways through which outputs of one system are imposed as inputs to another system.

Laws that govern physical phenomena, however, merely impose relations on the system variables, while interconnection means that variables are shared among subsystems. For

Digital Object Identifier 10.1109/MCS.2007.906923

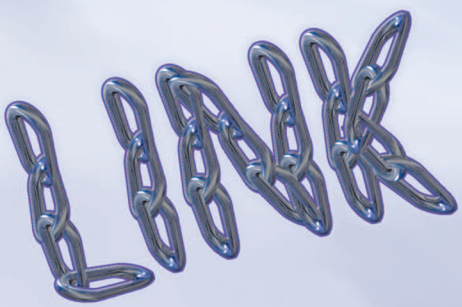
1066-033X/07/$25.00©2007IEEE

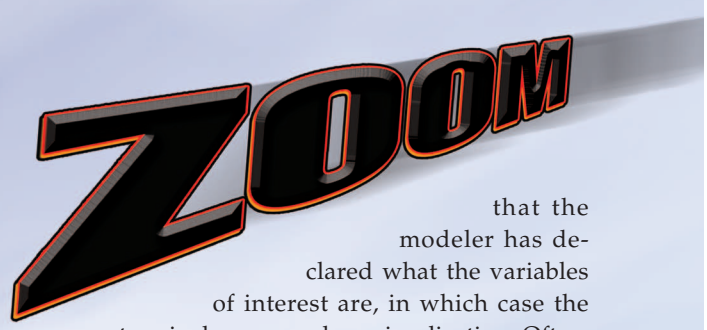
example, the gas law states how the variables of interest, temperature, volume, and mass are related. This law does not, however, state that some of the variables generate the others. The interconnection of two physical devices means that certain variables associated with the first device are set equal to certain variables associated with the second device. Connecting two pipes of two hydraulic systems means that the pressure and flow in the first pipe at the interconnection point are set equal to the pressure and flow in the second pipe at the interconnection point. After interconnection, the two hydraulic systems share the pressure and flow variables.

Relations as models of physical phenomena, as well as variable sharing to express interconnections, do not inherently involve signal flows. Viewing relations between system variables in terms of inputs and outputs, while viewing interconnection as output-to-input assignment, with signal transmission from one subsystem to another, usually introduces a signal transmission mechanism that is not part of the physics of the system or the interconnection. Signal-flow graphs are appropriate in some special, although important, situations, for example, in signal processing, in feedback control based on sensor outputs and actuator inputs, and in systems composed of unilateral devices. A unilateral device is a system that cannot be backdriven, such as an amplifier or a switch. But, as illustrated in this article, signal-flow diagrams are limited as a framework for dealing with mathematical descriptions of physical phenomena and with interconnections.

The notion of a behavior as a model treats all of the system variables on an equal footing. After analyzing the model, and depending on the purpose for which the model is used, it may be expedient to partition the system variables in two sets, input variables and output variables. The behavior provides a framework in which this input/output structure can be deduced. Classical input/output models are thus incorporated as behavioral models with additional structure. However, it is sometimes the case that input/output partitioning is impossible, and thus no separation of the system variables as inputs and outputs is possible.

A typical modeling task can be viewed as follows. The aim is to model the dynamic relations among several variables. We visualize this modeling problem by means of a black box with terminals (see Figure 1). One can think of these terminals as the places where these variables "live." In principle, the terminals and the black box express only





that the modeler has declared what the variables of interest are, in which case the terminals are merely a visualization. Often, though, the terminals are real, that is, physically available, and the aim is to model the variables associated with physical terminals through which a system can interact with its environment. When dealing with interconnections, it is natural to assume that these terminals and their variables are physical and to envision multiple physical variables collectively and indivisibly associated with a single terminal.

To fix ideas about the kind of situations and the nature of variables associated with terminals, it is helpful to think of the following examples, illustrated schematically in Figure 2.

» Forces and torques acting on the terminals of a mechanical structure as well as the displacements and attitudes of these terminals.

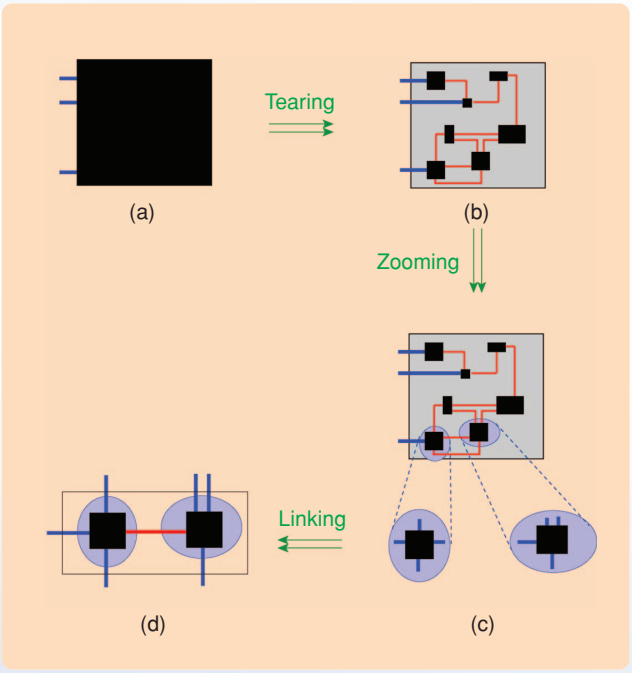


FIGURE 1 Modeling by tearing, zooming, and linking. Part (a) shows a black box with terminals. The aim is to obtain a model of the behavior of the variables on these external terminals. Part (b) shows the result of the tearing process: the black box is viewed as a gray box of interacting subsystems. The modeling process proceeds by zooming in on the subsystems one by one, as illustrated in (c). The subsystems are subsequently linked by sharing the variables on their common terminals, as illustrated by (d). The combination of the models of the subsystems and the interconnection constraints leads to a model of the variables on the external terminals. This modeling process has a hierarchical structure, since a subsystem can in turn be modeled by tearing, zooming, and linking.

» Currents and voltages associated with wires that connect an electrical circuit to its surroundings.
» Mass flows and pressures in pipes through which a fluid flows in and out of a hydraulic system.
» Heat flows and temperatures in ducts through which heat flows in and out of a thermodynamic engine, as well as mechanical work done by the engine.
» Actuator inputs and sensor outputs that interconnect a system with a controller.
» Combinations of the above, namely, multidomain devices, such as motors, pumps, and strain gauges, in which some of the terminals are mechanical, some electrical, some thermal, and some hydraulic or multidomain terminals that, for example, serve at the same time for heat conduction and mass transport, and are subject to mechanical forces.
» Restrictions of the above, for example, a mechanical system in which we are interested only in the displacements of the terminals or an electrical system in which we are interested only in the currents in the terminals, or a hydraulic system in which we want to model mass flow only; in other words, situations in which we are interested only in a subset of the physical variables associated with a terminal.
» Globalizations of the above, for example, a system in which we are interested in modeling only the energy that flows in and out of the system.

The black box in Figure 1 suggests that an underlying structure links the terminal variables and leads to the laws that govern them. Deriving these laws requires examining what is inside the black box. Systems often consist of interacting subsystems. To discover these interactions, we look inside the black box, where we find an interconnection architecture of smaller black boxes that interact through terminals of their own (see Figure 1). Modeling then proceeds by examining the smaller black boxes and their interactions. This modeling procedure is called *tearing*, *zooming*, and *linking*, in which we have the following:

1) *Tearing* refers to viewing a system as an interconnection of subsystems.
2) *Zooming* refers to modeling the subsystems.
3) *Linking* refers to modeling the interconnections among the subsystems.

This modeling process has an obvious hierarchical structure. Indeed, zooming involves modeling the laws that govern the variables on the terminals of a subsystem. This subsystem may in turn consist of interacting subsubsystems. Modeling the subsystem then again involves tearing, zooming, and linking. This process goes on until we

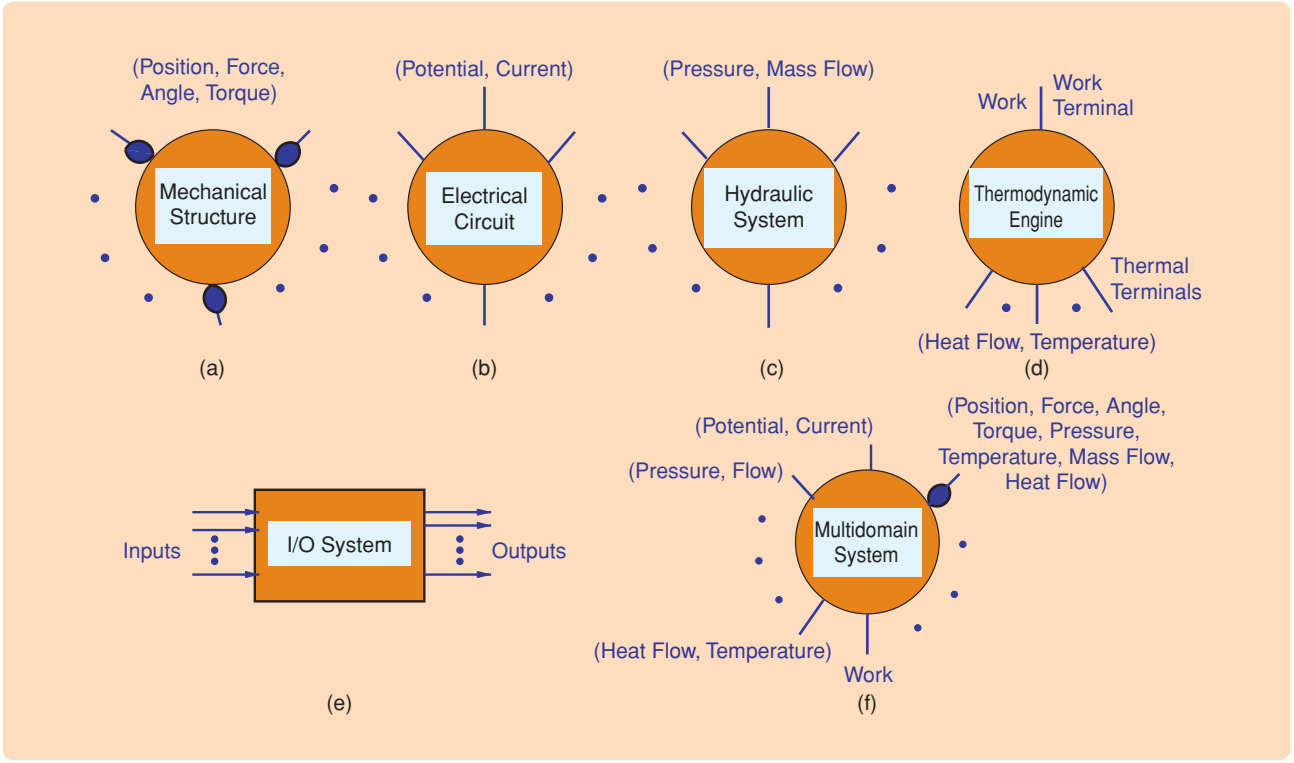


FIGURE 2 Examples of systems with terminals. (a) A mechanical system can be interconnected with its environment through terminals, each of which has a position and an attitude as well as a force and a torque acting on it. In the case of (b) an electrical system, the interaction takes place through wires, with a potential and a current associated with each wire. For (c) a hydraulic system, the terminals are outlets, each with an associated pressure and mass flow. For (d) a thermal terminal, these variables are temperature and heat flow. To express the second law of thermodynamics, a terminal can be used to visualize the work done on the environment. For (e) a system modeled with a signal flow, we have the usual input and output terminals. (f) Multidomain systems have different types of terminals.

meet a component whose model follows from first principles, a subsystem whose model has been stored in a database, or where system identification is the appropriate modeling procedure.

The purpose of this article is to develop a mathematical language for dealing with models of open systems and their interconnections. A mathematical framework that conceptualizes the dynamics of open systems must aim at the dynamics of the variables on the terminals, and it must also be capable of dealing with the interconnection constraints that result from connecting physical terminals of subsystems. The assertion is *that the behavioral approach provides a language that respects the physics*. Although input/output thinking is useful in certain situations, we argue throughout this article that, as a general methodology, input/output descriptions are ill-founded and clash with system interconnection. Interconnection, as we shall see, results in *variable sharing*, not in output-to-input assignment.

AN ILLUSTRATIVE EXAMPLE

In the next section, we introduce the notion of a dynamical system in terms of a behavior, while later in the article we formalize the general methodology of modeling interconnected systems by tearing, zooming, and linking. But before delving into these generalities, we consider an elementary example to motivate the ideas. This example is purely pedagogical and is accessible without the aid of any formalism whatsoever. The example is illustrative of more complex systems, where there is no real alternative to tearing, zooming, and linking.

Consider a hydraulic system consisting of two tanks filled with a fluid and connected by a pipe (see Figure 3). The system has two external outlets. We wish to model the relation between the variables p_{left}, f_{left}, p_{right}, f_{right}, which are the pressures and mass flows at these outlets. In other words, we wish to specify the possible time trajectories $(p_{\text{left}}, f_{\text{left}}, p_{\text{right}}, f_{\text{right}}) : \mathbb{R} \to \mathbb{R}^4$. This collection of time trajectories is what we mean by a dynamical model.

The procedure followed for modeling this physical system is illustrated in Figure 3. We view this system as a black box with two terminals, namely, the two outlets, and with two variables, namely, a pressure and a mass flow, associated with each of these terminals.

Tearing

In the tearing step, the system is viewed as an interconnection of subsystems. Looking inside the black box, we find three black boxes, two tanks (black boxes 1 and 3), and one pipe (black box 2).

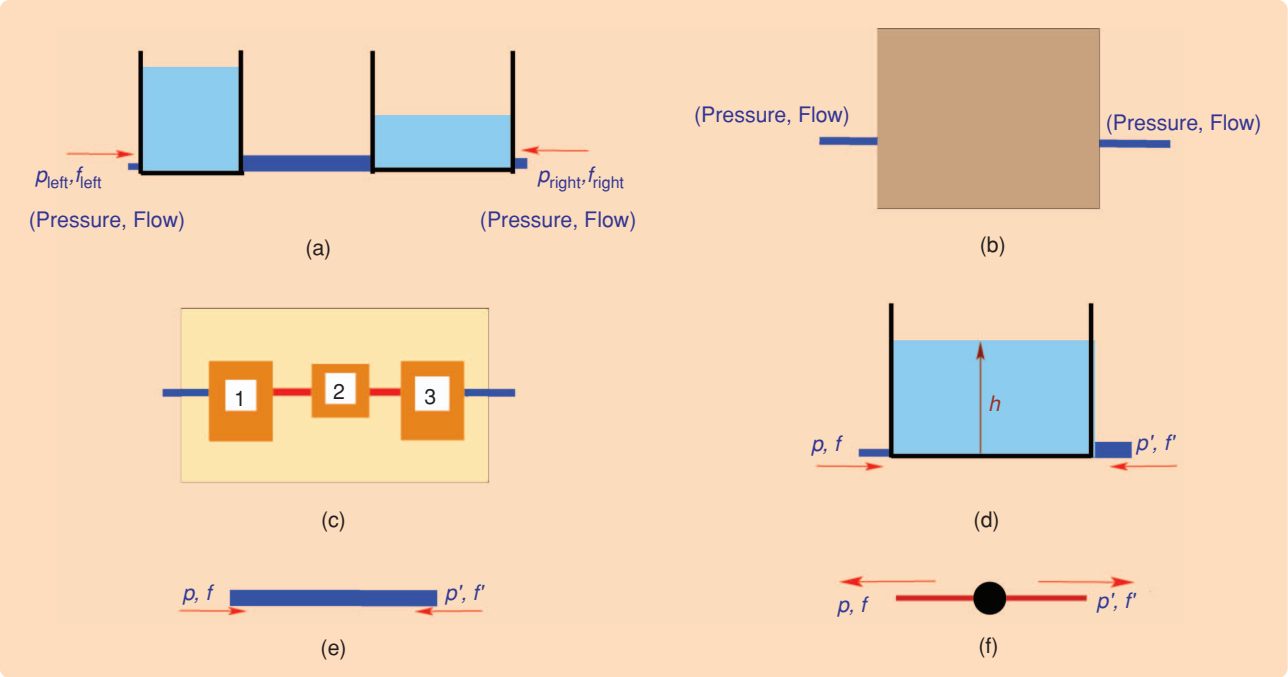


FIGURE 3 An example of modeling an interconnected system. Part (a) shows a hydraulic system consisting of two tanks connected by a pipe. The aim is to model the pressures and flows at the external outlets. This system is visualized in (b) as a black box with two external terminals, each with two associated variables, a pressure and a flow. The black box is viewed as (c) a gray box, consisting of an interconnection of three black boxes. These subsystems are modeled one by one. Subsystems 1 and 3, one of which is shown in (d), are simple tanks, while (e) subsystem 2 is a pipe. The relations between the terminal variables of the subsystems can be modeled from first principle physical laws. The interconnections, shown in (f), equate the pressures of the interconnected terminals at the interconnection points and put the sum of the flows equal to zero. Combining the subsystem models and interconnection constraints leads to a model of the variables at the external outlets. Interconnection is variable sharing. For the example at hand, inputs and outputs add an unphysical artifact in viewing the relation between the terminal variables, and in interpreting the interconnection of the subsystems.

Zooming

In the zooming step, we examine each subsystem individually and model the behavior of their terminal variables. We start with the black box 1, a hydraulic system with two outlets, each with two terminal variables, a pressure and a flow. We set out to discover the dynamic relations among these four variables. At this point, there is no more need for tearing, since the system in black box 1 is simple enough to be modeled using first principles physical laws. In a more complex application, the tearing and zooming processes could continue for several more layers. Let p, p' denote the pressures at the outlets and f, f' the mass flows, counted positive when fluid runs into the tank, as indicated by the arrow at the outlet in Figure 3. Of course, the dynamical laws depend on the material and geometric properties of the fluid and the tank, for example, the specific weight of the fluid, the surface area of the tank, and the cross sections of the outlets. We assume that the top of the tank is at constant atmospheric pressure, denoted by p_0. To determine the relations governing p, f, p', f', we introduce the height $h > 0$ of the fluid level in the tank as an auxiliary variable. Using the law of Daniel Bernoulli and conservation of mass leads to the dynamic equations

$$A\frac{d}{dt}h = f + f', \tag{1}$$

$$Bf = \begin{cases} \sqrt{|p - p_0 - \rho h|}, & \text{if } p - p_0 \geq \rho h, \\ -\sqrt{|p - p_0 - \rho h|}, & \text{if } p - p_0 \leq \rho h, \end{cases} \tag{2}$$

$$Cf' = \begin{cases} \sqrt{|p' - p_0 - \rho h|}, & \text{if } p' - p_0 \geq \rho h, \\ -\sqrt{|p' - p_0 - \rho h|}, & \text{if } p' - p_0 \leq \rho h, \end{cases} \tag{3}$$

where A, B, C, ρ, p_0 are physical constants that depend on the geometry and material properties.

Next, consider black box 3. For the case at hand, this system is similar to black box 1, except, perhaps, for the geometry. The equations describing its terminal variables are identical to those of the first tank, with possibly different parameters A, B, C.

Consider finally black box 2. This system is a pipe for transporting fluid. Let p, p' denote the pressures at the outlets and f, f' the mass flows. The notation f, p, f', p' is local and thus unrelated to the notation used in (1)–(3). The equations connecting these variables express the incompressibility of the fluid and the resistive relation between the pressure drop across the pipe and the mass flow rate through the pipe. A more accurate model could involve dynamics, hysteresis, and distributed effects. For simplicity, however, this relationship is taken to be linear and memoryless, leading to

$$f = -f', \quad p - p' = \alpha f, \tag{4}$$

where the constant $\alpha \geq 0$ depends on the geometry and material properties of the fluid and the pipe.

Linking

To obtain the complete equations of the interconnected hydraulic system, we also need to express the interconnection laws, that is, the relations that result from the fact that the terminals of the three black boxes are connected. Each of the interconnections involves two pressures and two mass flows, leading to the relations (note again the local nature of the notation)

$$f + f' = 0, \quad p = p'. \tag{5}$$

Setting up these interconnection constraints constitutes linking.

The Model Equations

The combination of the equations obtained by tearing, zooming, and linking leads—in the obvious notation—to the dynamic equations

$$A_1\frac{d}{dt}h_1 = f_1 + f_1', \tag{6}$$

$$B_1 f_1 = \begin{cases} \sqrt{|p_1 - p_0 - \rho h_1|}, & \text{if } p_1 - p_0 \geq \rho h_1, \\ -\sqrt{|p_1 - p_0 - \rho h_1|}, & \text{if } p_1 - p_0 \leq \rho h_1, \end{cases} \tag{7}$$

$$Cf_1' = \begin{cases} \sqrt{|p_1' - p_0 - \rho h_1|}, & \text{if } p_1' - p_0 \geq \rho h_1, \\ -\sqrt{|p_1' - p_0 - \rho h_1|}, & \text{if } p_1' - p_0 \leq \rho h_1, \end{cases} \tag{8}$$

$$f_2 = -f_2', \quad p_2 - p_2' = \alpha f_2, \tag{9}$$

$$A_3\frac{d}{dt}h_3 = f_3 + f_3', \tag{10}$$

$$Cf_3 = \begin{cases} \sqrt{|p_3 - p_0 - \rho h_3|}, & \text{if } p_3 - p_0 \geq \rho h_3, \\ -\sqrt{|p_3 - p_0 - \rho h_3|}, & \text{if } p_3 - p_0 \leq \rho h_3, \end{cases} \tag{11}$$

$$C_3 f_3' = \begin{cases} \sqrt{|p_3' - p_0 - \rho h_3|}, & \text{if } p_3' - p_0 \geq \rho h_3, \\ -\sqrt{|p_3' - p_0 - \rho h_3|}, & \text{if } p_3' - p_0 \leq \rho h_3, \end{cases} \tag{12}$$

$$p_1' = p_2, \quad f_1' + f_2 = 0, \tag{13}$$

$$p_2' = p_3, \quad f_2' + f_3 = 0, \tag{14}$$

$$p_{\text{left}} = p_1, \quad f_{\text{left}} = f_1, \quad p_{\text{right}} = p_3', \quad f_{\text{right}} = f_3'. \tag{15}$$

Equations (6)–(15) form a dynamic model relating the pressures and mass flows $p_{\text{left}}, f_{\text{left}}, p_{\text{right}}, f_{\text{right}}$, the four variables whose dynamic behavior we set out to model. Note that (6)–(15) involve the auxiliary variables $h_1, p_1, f_1, p_1', f_1', p_2, f_2, p_2', f_2', h_3, p_3, f_3, p_3', f_3'$ in addition to the variables of interest $p_{\text{left}}, f_{\text{left}}, p_{\text{right}}, f_{\text{right}}$.

It is illustrative to reflect on the following issues in the context of this example.

» This type of model is the end result of a systematic tearing, zooming, and linking procedure. We wish to

make an abstraction of this end result the beginning of a general theory of dynamics. What mathematical definition of dynamical system does this model suggest? What are the relevant concepts?

- » What do (6)–(15) express about p_{left}, f_{left}, p_{right}, f_{right}? In what sense do these equations define a dynamical system? When another modeler comes up with a different set of equations, when can we declare this new set of model equations equivalent to (6)–(15)?
- » Equations (6)–(15) involve the auxiliary variables $h_1, p_1, f_1, p_1', f_1', p_2, f_2, p_2', f_2', h_3, p_3, f_3, p_3', f_3'$ in addition to the variables p_{left}, f_{left}, p_{right}, f_{right}, which the model aims at. Can these auxiliary variables be eliminated? Note that some of these variables can be immediately eliminated, using (13)–(15). Is it possible to eliminate all of these auxiliary variables and obtain a set of differential equations involving only the variables p_{left}, f_{left}, p_{right}, f_{right}? Is this elimination a useful thing to do?
- » How would this modeling task evolve in an input/output mode of thinking? Which variables act as inputs and outputs of the interconnected hydraulic system described by the combined equations (6)–(15)? Which variables act as inputs and outputs for the subsystem described by (1)–(3) and for the subsystem described by (4)? Is it useful to think of these models and the resulting open systems in terms of inputs and outputs? Does it make sense to think of the interconnection equations (5), (13), and (14) as output-to-input assignments? If so, what would be the inputs and outputs?
- » Is this system controllable? Is controllability a valid question? Since this system is not in the classical input/state/output form, what would controllability mean?
- » Judging from this example, is an ordinary differential equation

$$f\left(w(t), \frac{d}{dt}w(t), \ldots, \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}}w(t)\right) = 0 \tag{16}$$

in the variables $w = (p_{\text{left}}, f_{\text{left}}, p_{\text{right}}, f_{\text{right}})$, which the model aims at, a good starting point for a general model description for a theory of nonlinear differential dynamical systems? Is a state model

$$\frac{d}{dt}x(t) = f(x(t), u(t)), \quad y(t) = h(x(t), u(t)), \tag{17}$$

with, for the case at hand, (u, y) equal, up to reordering of the components, to $(p_{\text{left}}, f_{\text{left}}, p_{\text{right}}, f_{\text{right}})$ and $x = (h_1, h_3)$, a better starting point? Or do the model equations (6)–(15) suggest a more general starting point that has both (16) and (17) as special cases?

These issues, as well as related questions, are dealt with in a general setting in this article.

THE BASIC CONCEPTS

In this section, we introduce the basic concepts of the behavioral language for modeling dynamical systems. The first questions that need to be confronted are: *What are we after? When we accept a mathematical model as a description of a phenomenon, what do we really assume? When we declare a set of equations to be a mathematical model, what does this statement mean?* These questions are not meant to be philosophical but mathematical. The answer to these questions is simple, evident, but enlightening and pedagogically effective.

The behavioral framework views a model as follows. Assume that we have a phenomenon that we wish to describe mathematically. Nature, that is, the reality that governs the phenomenon, can produce certain *events*, also called *outcomes*. The totality of feasible events, *before* we have specified laws that govern the phenomenon, forms a set $\mathbb{V}$, the universum of feasible outcomes. A *mathematical model* of the phenomenon restricts the outcomes that are declared possible to a subset $\mathcal{B}$ of $\mathbb{V}$; $\mathcal{B}$ is the *behavior* of the model. $(\mathbb{V}, \mathcal{B})$, or the subset $\mathcal{B}$ by itself, since $\mathbb{V}$ is usually evident from the context, is what we consider to be a mathematical model.

To illustrate this elementary idea, consider the ideal gas law, which poses $PV = RNT$ as the relation between the pressure P, volume V, mass N (the number of moles), and temperature T of an ideal gas, with R a constant that is the same for all gases. The universum $\mathbb{V}$ is $(0, \infty)^4$ and the behavior $\mathcal{B}$ is $\{(P, V, N, T) \in (0, \infty)^4 | PV = RNT\}$.

In the study of dynamical systems, we are interested in situations where the events are maps from a set of time instances to a set of outcomes. The universum is then the collection of all maps from the set of independent variables to the set of dependent variables. In models of physical phenomena, it is customary to call the elements of the domain of a map independent variables and those of the codomain dependent variables. For dynamical systems, the independent variable is time, and the set of independent variables is therefore a subset of $\mathbb{R}$. Later in the article, we discuss spatially distributed systems described by partial differential equations (PDEs), which involve multiple independent variables, reflecting, for example, time and space. But now, we discuss only dynamical systems where $\mathbb{T}$ is a set of real numbers. The set of dependent variables $\mathbb{W}$ is the set in which the outcomes of the signals being modeled take on their values. We call $\mathbb{T}$ the *time axis* and $\mathbb{W}$ the *signal space*. Hence a *dynamical system* is defined as a triple

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

with the behavior $\mathcal{B}$ a subset of $\mathbb{W}^{\mathbb{T}}$, where $\mathbb{W}^{\mathbb{T}}$ denotes the set of all maps from $\mathbb{T}$ to $\mathbb{W}$. For a dynamical system, the

universum of all possible events is hence $\mathbb{V} = \mathbb{W}^{\mathbb{T}}$, and the behavior $\mathcal{B}$ is a subset of it. Of course, for continuous-time systems, behaviors $\mathcal{B}$ of interest consists of a strict subset of $\mathbb{W}^{\mathbb{T}}$. In applications, elements of $\mathcal{B}$ are required to be well-behaved maps from $\mathbb{T}$ to $\mathbb{W}$, at least measurable or locally integrable. In fact, when studying linear time-invariant differential system, we often assume for convenience of exposition that the elements of $\mathcal{B}$ are infinitely differentiable.

The behavior $\mathcal{B}$ is the central object in this definition. The behavior formalizes which trajectories $w : \mathbb{T} \to \mathbb{W}$ are possible, according to the model. In the sequel, the terms "dynamical model," "dynamical system," and "behavior" are used as synonyms, since usually $\mathbb{W}$ and $\mathbb{T}$ follow from the context, leaving only $\mathcal{B}$ as being specified by the model equations.

As an example, consider the motion of a planet around the sun. For this example, the time axis is $\mathbb{R}$ and the signal space is $\mathbb{R}^3$, since we are interested in describing the position trajectories that the planet can trace out. Before these motions were understood, every trajectory $w : \mathbb{R} \to \mathbb{R}^3$ could conceivably occur. Kepler's laws limit the behavior to trajectories (K1) that are ellipses with the sun located at one of the foci, (K2) for which the line segment from the sun to the planet sweeps out equal areas in equal time, and (K3) such that the square of the major axis divided by the third power of the period of revolution is equal to a constant that is the same for all planets. So the behavior $\mathcal{B}$ articulated by Kepler's three laws is a small, but well defined, subset of $(\mathbb{R}^3)^{\mathbb{R}}$.

The behavioral framework treats a model for what a model ought to be, namely, an exclusion law.

The Behavior Is All There Is

Equivalence of models, representations of models, properties of models, and approximation of models must all refer to the behavior. The operations allowed to bring model equations in a more convenient form are exactly those that do not change the behavior. Dynamical modeling and system identification aim at coming up with a specification of the behavior. Control comes down to restricting the behavior. Expositions of this approach to system theory as presented here are given in [1]–[4], an early source is [5]. But, since this point of view is so natural, similar ideas can be found in other domains, for example, electrical circuit theory [6], [7], general systems theory [8], [9], and the theory of automata and machines [10]. The aim of this article is to motivate and explain the ideas of the behavioral approach. A summary of the main issues covered is given in "The Behavioral Approach."

We illustrate the use of the behavior to formulate system-theoretic concepts by means of two often used properties of dynamical systems, namely, linearity and time invariance. The dynamical system $\Sigma = (\mathbb{T}, \mathbb{W}, \mathcal{B})$ is *linear* if $\mathbb{W}$ is a vector space and $\mathcal{B}$ a linear subspace of $\mathbb{W}^{\mathbb{T}}$, that is, if $w_1, w_2 \in \mathcal{B}$ implies $\alpha w_1 + \beta w_2 \in \mathcal{B}$ for all scalars α, β. Linearity means that superposition and scaling hold. The dynamical system $\Sigma = (\mathbb{T}, \mathbb{W}, \mathcal{B})$ is *time invariant* if $\mathbb{T}$ is closed under addition and $\sigma^t\mathcal{B} \subseteq \mathcal{B}$ for all $t \in \mathbb{T}$, where σ^t denotes the backward t-shift, defined by $(\sigma^t f)(t') := f(t' + t)$. Time invariance means that the shift of a legal trajectory is again legal.

With a *representation* of a behavior we mean a formula, an expression, or a rule that specifies which elements of $\mathbb{W}^{\mathbb{T}}$ belong and which elements of $\mathbb{W}^{\mathbb{T}}$ do not belong to the behavior. Representations of the same behavior can look very different. For dynamical systems, we are used to thinking of a model as a set of differential equations. However, not all dynamical systems come as differential equations. Kepler's laws are a nice example of a description that directly spells out the trajectories in the behavior of a dynamical system, without intervention of differential equations. The bounded solutions of the second-order differential equation that emerges from application of Newton's laws to the motion of the planets is a representation of Kepler's laws.

Latent Variables

In applications, the behavior $\mathcal{B}$ must be specified, and it is here that differential or difference equations as well as alternative system representations enter the scene. An additional element, which enters models and modeling from the beginning, is *latent variables*. Latent variables are auxiliary variables that are involved in a model but that are not the variables the model aims at. Latent variables are ubiquitous in models, but, at first sight, they may seem perhaps elusive and superfluous. Therefore, we first discuss some examples and then turn to formal definitions.

For the system discussed in the section "An Illustrative Example," the aim is to set up a model for the behavior of the variables p_{left}, f_{left}, p_{right}, f_{right}. However, in arriving at the model equations (6)–(15), we found it necessary to introduce the auxiliary variables $h_1, p_1, f_1, p_1', f_1'$, $p_2, f_2, p_2', f_2', h_3, p_3, f_3, p_3', f_3'$. Note that even in modeling the relation among the variables p, f, p', f' of the simple tank in black box 1 and 3, leading to (1)–(3), we found it convenient to introduce the height h as an auxiliary variable. State variables are examples of the usefulness of latent variables in general dynamical models. Input/state/output models, such as (17), have the special feature that they specify the relation between inputs and outputs through a set of auxiliary variables, the state variables.

Great flexibility is obtained by expressing a model with the aid of auxiliary variables. We therefore incorporate these variables firmly in the behavioral modeling language and distinguish between the variables that the model aims at and the auxiliary variables introduced in the modeling process. The former are the *manifest* variables, while the latter are the *latent* variables. The use of latent variables, as state variables to express the relation between inputs and outputs, leads to models that are closer to physics, have

much more modeling power, and are easier to obtain and analyze than models that contain only inputs and outputs, such as Volterra series. Examples of latent variables in physics include the potentials in Maxwell's equations, discussed in the section "PDEs."

A *dynamical system with latent variables* is defined as

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

where $\mathbb{T}$ is the time axis, $\mathbb{W}$ is the set of manifest variables, $\mathbb{L}$ is the set of latent variables, and $\mathcal{B}_{\text{full}} \subseteq (\mathbb{W} \times \mathbb{L})^{\mathbb{T}}$ is the *full behavior*. The system Σ_{full} *induces*, or *represents*, the *manifest dynamical system* $\Sigma = (\mathbb{T}, \mathbb{W}, \mathcal{B})$, with the *manifest behavior* $\mathcal{B}$ defined by

$$\mathcal{B} = \{w : \mathbb{T} \to \mathbb{W} \mid \text{there exists } \ell : \mathbb{T} \to \mathbb{L} \text{ such that } (w, \ell) \in \mathcal{B}_{\text{full}}\}.$$

Latent variables are ubiquitous in mathematical models, witness the relevance of, for example, state variables in state models, interconnection variables in models obtained by tearing, zooming, and linking, potentials in Maxwell's equations, driving variables in image representations, the basic probability space in stochastics, the wave function in quantum mechanics, and the entropy in thermodynamics. A system with latent variables is the natural endpoint of first principles modeling, and hence the natural starting point for the analysis and synthesis of systems. Latent variables also enter forcefully in representation questions.

Behavioral Equations

In many applications, the behavior $\mathcal{B}$ is given through a system of equations, which, for continuous-time dynamical systems, is usually a system of differential equations. The objective of this section is to determine a suitably general class of differential equations as a starting point for the study of dynamics. Is an explicit differential equation in the manifest variables, as (16), with the manifest variables $w = (u, y)$ partitioned in inputs and outputs, a reasonable starting point? Or is the state-space version, as

The Behavioral Approach

The behavioral approach is based on the following premises.

1) A mathematical model is a subset of a set of a priori possibilities. This subset is the behavior of the model. For a dynamical system, the behavior consists of the time trajectories that the model declares possible.
2) The behavior is often given as a set of solutions of equations. Differential and difference equations are an effective, but highly nonunique, way of specifying the behavior of a dynamical system.
3) The behavior is the central concept in modeling. Equivalence of models, properties of models, model representations, and system identification must refer to the behavior.
4) Both first principles models and models of interconnected systems usually contain latent variables in addition to the manifest variables that the model aims at. Elimination of latent variables compactifies the behavioral equations. For linear time-invariant differential systems, complete elimination of latent variables is possible.
5) Physical systems are usually not endowed with a signal flow graph. Input/output models of physical systems are appropriate only in some special situations.
6) Interconnected systems can be modeled using tearing, zooming, and linking. The interconnection architecture can be formalized as a graph with leaves. The nodes of the graph correspond to the subsystems, the edges correspond to the connected terminals, and the leaves correspond to terminals by which the interconnected system interacts with its environment. Interconnection of physical systems means variable sharing. Output-to-input assignment is often an unnecessary, inconvenient, and limiting way of viewing physical interconnections.
7) System-theoretic concepts such as controllability and observability are simpler to define and more general in the behavioral setting than in the state-space setting. Controllability becomes a genuine property of a dynamical system rather than of just a state representation.
8) Control means restricting the behavior of a plant by interconnection with a controller. Control by input selection, that is, open-loop control, and by feedback, that is, closed-loop control, are special cases.
9) Linear time-invariant differential systems (including the special case of differential-algebraic systems) are in one-to-one correspondence with $\mathbb{R}[\xi]$ submodules. This correspondence provides the ability to translate every property of a linear time-invariant differential behavior into a property of the associated submodule. Since these $\mathbb{R}[\xi]$ submodules are finitely generated, computer-algebra-based algorithms can be used to analyze the system properties.
10) For linear time-invariant differential systems, controllability is equivalent to the existence of an image representation, as well as to the case that the corresponding $\mathbb{R}[\xi]$ module is closed. Controllable linear time-invariant differential systems are in one-to-one correspondence with $\mathbb{R}(\xi)$-subspaces.
11) One-to-one correspondence of linear time-invariant systems with submodules, elimination of latent variables, and equivalence of controllability with the existence of an image representation are also valid for systems defined by constant-coefficient linear partial differential equations.

(17), with the latent variables restricted to state variables, a better one?

Assume that the time set is $\mathbb{R}$, and that the signal space $\mathbb{W}$ has enough structure for differentiation to be defined. The behavior defined by (16) is then

$$\mathcal{B} = \left\{ w : \mathbb{R} \to \mathbb{W} \mid f\left(w(t), \frac{d}{dt}w(t), \dots, \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}}w(t)\right) = 0 \text{ for all } t \in \mathbb{R} \right\}.$$

Of course, this definition requires an appropriate solution concept for differential equations. We gloss over this issue for now. When latent variables are present, we obtain instead

$$f\left(w(t), \frac{d}{dt}w(t), \dots, \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}}w(t), \ell(t), \frac{d}{dt}\ell(t), \dots, \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}}\ell(t)\right) = 0. \tag{18}$$

The full behavior $\mathcal{B}_{\text{full}}$ is then

$$\mathcal{B}_{\text{full}} = \left\{ (w, \ell) : \mathbb{R} \to \mathbb{W} \times \mathbb{L} \mid f\left(w(t), \frac{d}{dt}w(t), \dots, \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}}w(t), \ell(t), \frac{d}{dt}\ell(t), \dots, \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}}\ell(t)\right) = 0 \text{ for all } t \in \mathbb{R} \right\}. \tag{19}$$

Obviously, a state-space model (17) is a special case of (18), with the state a latent variable. Systems in which the manifest behavior or the full behavior is defined as the solution set of a system of differential equations, such as (16) or (18), are called *differential systems*.

The Elimination Problem

One question that emerges is whether the manifest behavior of a differential system with latent variables is itself also a differential system. Explicitly, the question is whether the manifest behavior corresponding to (19), that is,

$$\mathcal{B} = \left\{ w : \mathbb{R} \to \mathbb{W} \mid \text{ there exists } \ell : \mathbb{R} \to \mathbb{L} \text{ such that } f\left(w(t), \frac{d}{dt}w(t), \dots, \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}}w(t), \ell(t), \frac{d}{dt}\ell(t), \dots, \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}}\ell(t)\right) = 0 \text{ for all } t \in \mathbb{R} \right\}$$

is also the solution set of a system differential equations of the form

$$f'\left(w(t), \frac{d}{dt}w(t), \dots, \frac{d^{\mathtt{n}'}}{dt^{\mathtt{n}'}}w(t)\right) = 0$$

for a suitable f'. In effect, this question asks whether the solution set of a system of differential equations is closed under projection, that is, if latent variables can be eliminated from (19). Elimination theory and the associated algorithms is a much studied topic in mathematics, in particular in algebraic geometry [11].

As an example of elimination of latent variables, consider (1)–(3), and assume, for simplicity, that $A = B = C = \rho = 1$. It is easy to see that the time functions p, f, p', f', for which there exists a time function h such that (1)–(3) are satisfied, are exactly those that satisfy the differential-algebraic equations

$$p - p' = f|f| - f'|f'|, \frac{d}{dt}p - |f|\frac{d}{dt}f = f + f',$$

or, equivalently,

$$p' - p = f'|f'| - f|f|, \frac{d}{dt}p' - |f'|\frac{d}{dt}f' = f' + f.$$

These equations no longer contain h. In this example, elimination of h is indeed possible. Elimination of $h_1, p_1, f_1, p_1', f_1', p_2, f_2, p_2', f_2', h_3, p_3, f_3, p_3', f_3'$ is also possible for the full equations (6)–(15), leading to differential-algebraic equations involving only p_{left}, f_{left}, p_{right}, f_{right}.

It is easy to construct simple examples of differential equations for which elimination fails. For example, elimination of ℓ from the differential equation $\ell(t)(d/dt)w(t) = 1$ for all $t \in \mathbb{R}$, leads to the differential inequation $(d/dt)w(t) \neq 0$ for all $t \in \mathbb{R}$ for the manifest behavior. Likewise, $|(d/dt)w|^2(t) + |(d/dt)\ell|^2(t) = 1$ for all $t \in \mathbb{R}$, leads to the differential inequality $|(d/dt)w(t)| \leq 1$ for all $t \in \mathbb{R}$ for the manifest behavior. In both examples, the full behavior with latent variables is described by a differential equation, but the manifest behavior is not. The view that the manifest behavior of a smooth nonlinear dynamical system can be described as the solution set of a system of differential equations is classical. But, as we have just seen, because of the elimination problem, this assumption is much less innocent for nonlinear systems than it appears. When and how it is possible to eliminate latent variables from nonlinear differential systems and remain in the class of systems described by a differential equation is a complicated matter. That the behavior of smooth nonlinear dynamical system can be described as the solution set of a system of differential equations in the system variables does not follow in a straightforward way from the assumption that the subsystems are smooth and described by differential equations.

Image representations form another class of systems that have come to play a role in system theory. In this case, the model equations are of the form

$$w(t) = f\left(\ell(t), \frac{d}{dt}\ell(t), \dots, \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}}\ell(t)\right), \tag{20}$$

which is obviously a special case of (18). Note that (20) leaves the latent variable ℓ unconstrained, and hence the

manifest behavior is the image of the map f, whence the terminology "image representation." Behaviors expressed as an image representation are convenient in simulation, since it suffices to choose an arbitrary function ℓ to obtain a typical element w of the behavior. Dynamical systems that allow an image representation are *differentially flat* [12]. Not all behaviors allow such a representation. Whether a behavior admits an image representation is again not a matter of smoothness, but related to controllability, as discussed in the section "Controllability and Image Representations."

INPUTS AND OUTPUTS

Viewing the interaction of a system with its environment in an input/output manner is intuitively appealing. Inputs and outputs connote action and reaction. The environment acts by imposing certain variables, the inputs, on the system, while the system reacts by imposing certain variables, the outputs, on the environment. We thus arrive at the black box shown in figures 2 and 5.

This input/output view is eminently suitable in numerous situations. For example, inputs and outputs can serve to describe the reactions of humans and animals to stimuli, to design intelligent devices, such as computers and signal processors, to respond to external commands, or to explain algorithms. However, as we demonstrate in this section, viewing physical systems in terms of inputs and outputs is often a deficient way of expressing a dynamical model. Mathematical models state the simultaneous occurrence of physical variables, not that one variable causes another. In addition, viewing system interconnections in terms of inputs and outputs is often inappropriate in physical applications.

A map induces, through its graph, a relation on the Cartesian product of its domain and codomain. The *graph* of $F : \mathbb{U} \to \mathbb{Y}$ is given by $\text{graph}(F) = \{(u, y) \in \mathbb{U} \times \mathbb{Y} | \ y = F(u)\}$. For mathematical models, the map $F : \mathbb{U} \to \mathbb{Y}$ induces the model $(\mathbb{U} \times \mathbb{Y}, \text{graph}(F))$ with behavior $\mathcal{B} = \text{graph}(F)$, which can be interpreted in terms of the input u and the output y, as the cause/effect map $y = F(u)$.

It is often possible to interpret a mathematical model of a physical phenomenon as a graph. For example, the ideal gas law can be viewed as a map that specifies the volume as a function of the pressure, temperature, and number of moles, leading to $V = RNTP^{-1}$, with output V. Similarly, we can arrive at $T = R^{-1}N^{-1}PV$, with output T. But neither of these representations expresses the idea behind the physics of the gas law. Universally interpreting physical models as input/output maps is also impractical, especially when we plan to store the model in a database with the aim to embed this model in a signal-flow graph when the model is needed.

The appropriateness of input/output models is not a philosophical issue but rather is about the use of proper mathematical notions: *Should we think of a mathematical model as a subset, as a relation, or as a map?* Mathematical models deal with the simultaneous occurrence of events, not with causation. Systems interact by sharing variables, and it is not clear which variable is imposed by one subsystem on another subsystem. Does a driver impose a torque on the steering wheel and a force on the gas pedal, or is it an angle and a position that are imposed? The fact that these simple examples lead to dialectical dilemmas shows the weakness of input/output thinking. It is input/output thinking that is enamored with frivolous philosophy, by *ab initio* dragging in cause and effect, a slippery red herring of classical philosophy (see "Cause and Effect").

Another example in which input/output thinking leads to awkward situations is the ideal diode (see Figure 4). The current/voltage characteristic of an ideal diode is neither the graph of an impedance $I \mapsto V$ nor of an admittance $V \mapsto I$. We can view an ideal diode as an input/output system by taking the scattering variables $u = I + V$ and $y = I - V$ as input and output, but in models of diodes with a more complex voltage/current characteristic, this solution may not be possible. In more complex electrical devices, such as the series connection of two nonlinear resistors with a non-monotone characteristic (see Figure 4), it may not be possible to view the feasible voltage/current pairs of the series connection as the graph of a function. Additional examples showing the awkwardness of input/output thinking are resistors, gears, transformers, and two-sided devices, such as transmission lines and heat conduction bars.

Partitioning the System Variables into Inputs and Outputs

These arguments carry over to dynamical systems. A dynamical system is almost never an input/output map, since the response invariably also depends on the initial conditions. Incorporating initial conditions in system models is one of the merits of the state-space approach. However, in the context of models for open physical systems, input/state/output models require a partitioning, $w = (u, y)$, *from the very beginning*, of the variables w that the model aims at, into inputs u and outputs y. This partitioning cannot be made before we have studied the specific system but must be deduced from the concrete structure of the system, and hence this partition has to be based on a higher level description of the system. The behavior offers such a higher level description. Input/output descriptions, when deduced logically, therefore require behavioral models from which to deduce the input/output structure.

It is interesting to interpret some results from linear electrical circuit theory from this perspective; these issues apply as well to other domains, such as mechanics, hydraulics, and thermal systems. Consider a passive electrical circuit, containing linear positive resistors, capacitors, inductors, transformers, and gyrators, interconnected

in the usual way. Assume that there are wires, the external terminals through which the circuit can interact with its environment. Let V be the vector of external terminal potentials and I the vector of external terminal currents. The behavior of the variables $w = (V, I)$ is linear, time invariant, and differential. It can be proven, in other words, it is a theorem, that for such passive linear circuits, half of the external variables $w = (V, I)$ are inputs, while the other half of the external variables are outputs. Further, there is a choice of the input/output partition of $w = (V, I)$ such that each terminal contains exactly one input variable and exactly one output variable. Specifically, at each terminal, either the potential is an input and the current is an output, or the other way around, the current is an input and the potential is an output. Finally, one can always choose $V + I$ as the input, and $V - I$ as the output. These input/output representability results can only be theorems if there is higher level description from which to start. Behaviors provide such a higher level definition.

For some model classes, for example, for linear time-invariant differential systems, there always exists a componentwise input/output partition of the system variables. Consider the differential equation

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

where the real matrices $R_0, R_1, \ldots, R_{\mathtt{n}}$ are the parameters of the model, and the differential equation specifies which time trajectories $w : \mathbb{R} \to \mathbb{R}^{\mathtt{w}}$ belong to the behavior. This

Cause and Effect

Causality is thought and taught to be one of the pillars of system theory. The two most important properties of a cause-effect relation are

1) the cause leads to the effect
2) time wise, the cause precedes the effect.

Viewing a dynamical system as a nonanticipating input/output map captures both features very well, with the input the cause and the output the effect. Thus a dynamical system is often defined as a nonanticipating map F from an input space $\mathcal{U} \subseteq \mathbb{U}^{\mathbb{T}}$ to an output space $\mathcal{Y} \subseteq \mathbb{Y}^{\mathbb{T}}$, where $\mathbb{T} \subseteq \mathbb{R}$ denotes the time set, $\mathbb{U}$ is the set where the inputs take on their values, and $\mathbb{Y}$ is the set where the outputs take on their values; $\mathbb{U}^{\mathbb{T}}$ denotes the set of maps from $\mathbb{T}$ to $\mathbb{U}$, and $\mathbb{Y}^{\mathbb{T}}$ is similarly defined. The map F is *nonanticipating* if, for all $u_1, u_2 \in \mathcal{U}$ and $t \in \mathbb{T}$, the equality

$$u_1(t') = u_2(t') \quad \text{for all} \quad t' < t$$

implies the equality

$$y_1(t') = y_2(t') \quad \text{for all} \quad t' < t,$$

where $y_1 = F(u_1)$ and $y_2 = F(u_2)$. If $\mathcal{U}$ and $\mathcal{Y}$ are vector spaces and F is a linear map, then we arrive at the definition of a linear system used in many textbooks [S1] and in Wikipedia [S2].

Unfortunately, this definition of a dynamical system works only in the simplest examples. For example, the scalar differential equation $p(d/dt)y = q(d/dt)u$, where $p, q \in \mathbb{R}[\xi]$, can be thought of as describing a single-input/single-output linear system in this map sense, as does the feedback gain $y = -Ku$. But when this feedback is applied to the plant, we arrive at $(p(d/dt) + Kq(d/dt))y = 0$, and, suddenly, we seem to have left the realm of linear systems theory by the (feed-)back door. Indeed, $(p(d/dt) + Kq(d/dt))y = 0$ has no inputs, and hence there is no input/output map, but this system is, or ought to be, a bona fide linear system. The map definition also does not do justice to input/state/output systems, which model many more things by incorporating the fact that outputs also depend on initial conditions in addition to inputs.

What is a good mathematical definition of a cause-effect relation? Of nonanticipation? Consider the dynamical system $\Sigma = (\mathbb{T}, \mathbb{W}_1 \times \mathbb{W}_2, \mathcal{B})$, with $\mathcal{B}$ the behavior, consisting of pairs of trajectories (w_1, w_2), with $w_1 : \mathbb{T} \to \mathbb{W}_1$ and $w_2 : \mathbb{T} \to \mathbb{W}_2$. What do we mean by the statement that this model expresses that w_2 does not anticipate w_1? A logical way to proceed is as follows. First define the behavior consisting of the w_1-trajectories that the system declares possible. The system that governs w_1 is $\Sigma_1 = (\mathbb{T}, \mathbb{W}_1, \mathcal{B}_1)$, with $\mathcal{B}_1$ the projection of $\mathcal{B}$ onto $\mathbb{W}_1^{\mathbb{T}}$, that is,

$$\mathcal{B}_1 = \{w_1 : \mathbb{T} \to \mathbb{W}_1 \mid \text{there exists } w_2 : \mathbb{T} \to \mathbb{W}_2 \text{ such that } (w_1, w_2) \in \mathcal{B}\}.$$

We say that the outcomes of the signal w_2 *do not anticipate* the outcomes of the signal w_1 in the dynamical system $\mathcal{B}$ if, for all $w_1', w_1'' \in \mathcal{B}_1$, $t \in \mathbb{T}$, and $w_2' : \mathbb{T} \to \mathbb{W}_2$,

$$w_1'(t') = w_1''(t') \quad \text{for all} \quad t' < t, \quad \text{and} \quad (w_1', w_2') \in \mathcal{B}$$

imply that there exists $w_2'' : \mathbb{T} \to \mathbb{W}_2$ such that

$$(w_1'', w_2'') \in \mathcal{B} \quad \text{and} \quad w_2''(t') = w_2'(t') \quad \text{for all} \quad t' < t.$$

This definition expresses that, given the laws that govern the system, knowledge of the future of the trajectory w_1 in addition to its past does not give information about what could have happened in the past as far as the trajectory w_2 is concerned. In words, the definition of nonanticipation states that if the trajectories w_1' and w_1'' are equal in the past, differences among w_1' and w_1'' in the future cannot be detected by observing the associated trajectory w_2, since for any w_2' associated with w_1', meaning that $(w_1', w_2') \in \mathcal{B}$, there is a w_2'' associated with w_1'', meaning that $(w_1'', w_2'') \in \mathcal{B}$, that agrees with w_2' in the past.

Applied to systems described by linear time-invariant differential equations $R(d/dt)w = 0$, with $R \in \mathbb{R}[\xi]^{\bullet \times \bullet}$ (see the section "Linear Time-Invariant Differential Systems"), it

class of systems is studied in detail in the section "Linear Time-Invariant Differential Systems." It is easy to prove that in this case w can always be partitioned componentwise as $w \cong (u, y)$ ($\cong$ expresses the fact that equality holds only up to a reordering of the components) such that the behavior is exactly the set of solutions of

$$P_0 y + P_1 \frac{d}{dt} y + \cdots + P_{\mathtt{n}} \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}} y = Q_0 u + Q_1 \frac{d}{dt} u + \cdots + Q_{\mathtt{n}} \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}} u,$$

where $P(\xi) = P_0 + P_1 \xi + \cdots + P_{\mathtt{n}} \xi^{\mathtt{n}}$ and $Q(\xi) = Q_0 + Q_1 \xi + \cdots + Q_{\mathtt{n}} \xi^{\mathtt{n}}$ are polynomial matrices, with P square, $\det(P) \neq 0$, and $P^{-1}Q$ proper. These conditions on P and Q permit the interpretation of u as the free input and y as the output. Although the partition $w \cong (u, y)$ puts the input/output structure in evidence, this partition is not unique. In other words, for linear time-invariant differential systems, an input/output partition is always possible, componentwise, but this partition must be deduced from a behavioral model in which this input/output partition has not yet been made. In a sense, the existence of an input/output partition is unique to linear time-invariant differential systems. For time-varying systems, the input/output partition may depend on the time instance at which the partition is made, while, for nonlinear systems, the partition may depend on the operating point around which the partition is made. Just as a differentiable manifold is not necessarily the graph of a map, there is no reason to expect that a global input/output partition is possible for a smooth nonlinear dynamical system.

follows from the Smith form for polynomial matrices that each subset of the system variables is nonanticipating, in the sense defined above, with respect to every other subset of system variables. In other words, assume that the variables of the system are $w = (w_1, w_2, \ldots, w_{\mathtt{w}})$. Let $(w_1', w_2', \ldots, w_{\mathtt{m}}')$ and $(w_1'', w_2'', \ldots, w_{\mathtt{m}}'')$ be subsets of these variables. Then $(w_1', w_2', \ldots, w_{\mathtt{m}}')$ does not anticipate $(w_1'', w_2'', \ldots, w_{\mathtt{m}}'')$ in the behavior $R(d/dt)w = 0$. Consequently, this notion of nonanticipation does not distinguish variables in linear time-invariant differential systems due to the fact that differential equations impose laws that are local in time. Furthermore, nonanticipation forward in time implies nonanticipation backwards in time. It is hard to see how any definition of nonanticipation can distinguish the time direction in linear time-invariant differential systems. Nonanticipation becomes relevant, however, when considering discrete-time systems or systems with delays.

A continuous-time dynamical system described by constant-coefficient differential equations induces an input/output map, not because of some special structure of the dynamical laws that govern the system variables, but because initial conditions are imposed. For example, the behavior can become the graph of a nonanticipating input/output map by assuming zero initial conditions at $t = -\infty$, that is, by restricting the behavior to the system trajectories that have left compact support. But choosing initial conditions is an awkward thing to do. Initial conditions are not part of physical laws. For example, there is no reasonable way to choose universal initial conditions for the motion of a point mass in Newton's second law $F = (d^2/dt^2)q$ or for Maxwell's equations. The very idea of considering fixed initial conditions is awkward in nonlinear systems and therefore for linear systems that are interconnected with nonlinear ones. Basic properties of models of physical systems, such as nonanticipation and symmetry, are not concerned with initial conditions. Initialized systems are reasonable models in signal processing, where they incorporate the fact that messages start at some point and have finite duration. But the problem that nonanticipation does not distinguish variables or the time direction in linear time-invariant differential systems persists if we assume compact support solutions.

I do not want to be misunderstood. Obviously, cause and effect are part of our daily experience. Causality ought to be part of any theory that deals with unilateral phenomena and devices, such as amplifiers and switches. Defining interconnection architectures for systems that contain unilateral components is an important problem. Developing a mathematical language to deal with causality is a relevant issue. Often, causality is introduced in a stochastic context, but I do not find this approach convincing since I fail to understand what a probabilistic setting has to offer when the deterministic case has not yet been thought out properly.

It is remarkable that the idea of viewing a system in terms of inputs and outputs, in terms of cause and effect, kept its central place in systems and control theory throughout the 20th century. Input/output thinking is not an appropriate starting point in a field that has modeling of physical systems as one of its main concerns. Definitions of causality and nonanticipation suffer from the *post hoc ergo propter hoc* (it happened before, hence it caused) fallacy. Causality is one of those slippery red herrings of classical philosophy, witness the following quote of Bertrand Russell:

> *The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.* [S3]

REFERENCES

[S1] A.V. Oppenheim and A.S. Willsky, *Signals and Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[S2] Linear System, Wikipedia, [Online]. Available: http://en.wikipediaorg/wiki/Linear_system

[S3] B. Russell, "On the notion of cause," *Proc. Aristotelian Soc.*, vol. 13, pp. 1–26, 1913.

Difficulties with Input/Output Partitioning

There are also good mathematical reasons to be hesitant about imposing an input/output structure on the space of system variables. Some spaces may not allow a product space structure at all. Since partitioning a general set $\mathbb{W}$ as $\mathbb{W} = \mathbb{U} \times \mathbb{Y}$ is essentially a linear idea, or a local idea for smooth nonlinear systems, mathematical obstructions can impede a global input/output partition. Assume, for example (see Figure 4), that we wish to model the position and velocity of a moving body that travels freely on a manifold. This motion can be described by the model

$$\frac{d}{dt}y = u,$$

where the velocity u is the input and the position y is the output. However, for manifolds with a nontrivial tangent bundle, such as the unit sphere, this partition is locally, but not globally, possible.

Input/output partitioning is especially problematic in the context of interconnected systems. One of the premises implicit in system theory and implemented, for example, in Simulink, is that system interconnection can be viewed as output-to-input assignment. Let us confront this view with physical reality. First, observe that more than one physical variable is usually involved in interconnection constraints that result from the interconnection of two physical terminals. Connecting two electrical wires leads to the interconnection constraints $V_1 = V_2, I_1 + I_2 = 0$, where V_1 and V_2 are voltages, and I_1 and I_2 are currents. Likewise, soldering together two hydraulic pipes carrying a fluid leads to $p_1 = p_2, f_1 + f_2 = 0$, where p_1 and p_2 are pressures, and f_1 and f_2 are mass flows. Thermal connections lead to $T_1 = T_2, Q_1 + Q_2 = 0$, where T_1 and T_2 are temperatures, and Q_1 and Q_2 are heat flows. All of the terminals in these examples have more than one variable associated with them. If we insist on an input/output view, typically one of these variables acts as an input, and the other variable as an output. There is no reason that interconnection in these examples must correspond to output-to-input assignment (see Figure 5).

The universal classical picture of a system with an input terminal on one side and an output terminal on the other side, is pedagogically unfortunate for several reasons. To begin with, this picture shows two terminals for variables that often live on one and the same physical terminal. Further, this picture suggests that the input and output signals occur at different points, whereas these signals often act inseparably at the same physical point. Moreover, the signal-flow diagram suggests that the terminal to which the output variable at a particular terminal is directed can be different from the terminal from which the input variable of that particular terminal is obtained from (see Figure 5). This physical impossibility is not captured by signal-flow diagrams. By suggesting that these inputs and outputs act at different points, the input/output point of view fails to put physical reality in evidence.

In addition, physical interconnection constraints equate physical variables or equate their sum to zero, which is equating up to a sign. Voltages are identified with voltages, currents with currents, forces with forces, positions with positions, mass flows with mass flows, and pressures with pressures. Therefore, if physical intuition suggests that force is an input and position is an output, then interconnecting two mechanical systems by connecting two terminals leads to equating two inputs and equating two outputs, exactly the sort of connection that is forbidden in input/output thinking. The same holds for pressures as inputs and mass flows as

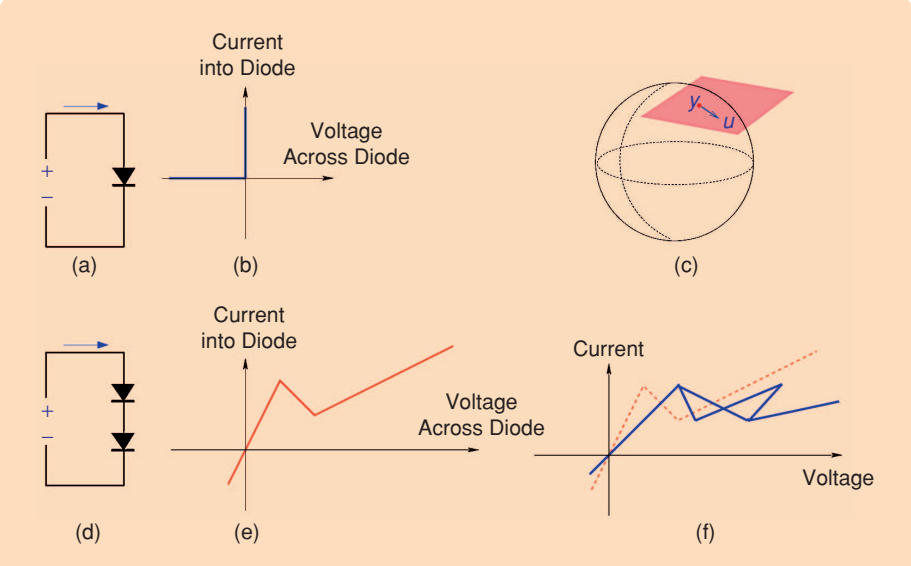


FIGURE 4 Systems that may not allow an input/output partition. These examples illustrate that input/output partition of terminal variables may be awkward. Part (a) shows an ideal diode, with (b) current/voltage port characteristic. Since an ideal diode is neither current driven nor voltage driven, the current/voltage port variables do not allow an input/output partition. Part (c) deals with the mathematical representation of the free motion of a point mass on a sphere. It is natural to expect the velocity to be the input and the position to be the output. However, because of the geometry of the tangent bundle of the sphere, such a partition is locally, but not globally, possible. The series connection (d) of two tunnel diodes with identical, nonmonotone, characteristic (e) leads to the feasible voltage/current pairs shown in (f). There is no convenient way in which this series connection can be viewed as an input/output system. In [13] series connections of tunnel diodes leading to disconnected voltage/current characteristics are presented.

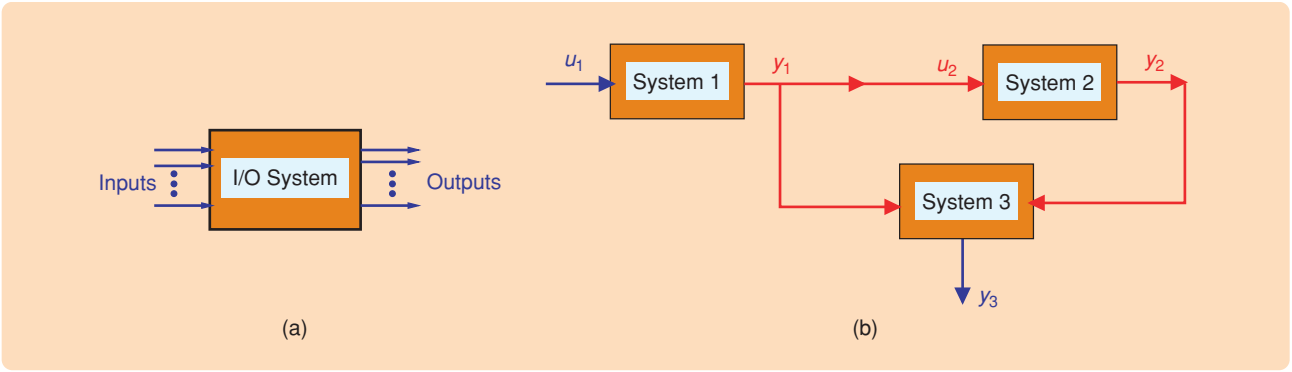


FIGURE 5 Input/output connections. Signal-flow graphs allow the generation of complex systems through interconnection architectures involving input/output systems (a) as building blocks and output-to-input assignment for interconnections. An example combining series and feedback connection is shown in (b). An input and an output often correspond to a partition of variables on a single physical terminal. When such an input/output partition is made, it is impossible to assign the output to be the input of another system without making the corresponding assignment of the input. For example, if the input and output of system 2 correspond to variables associated with a single physical terminal, then one cannot assign the output of system 1 to be the input of system 2 and at the same time assign the output of system 2 to be the input of system 3. The input and output of system 2 must both be directed to the same system when the interconnection corresponds to an interconnection of physical terminals. Signal-flow diagrams give a misleading view of how physical systems can be interconnected. In particular, interconnection architectures that do not respect the nature of interconnection of physical terminals are thus physically impossible, although they are allowed by signal-flow diagrams.

outputs in hydraulic systems as well as for temperatures as inputs and heat flows as outputs in thermal systems. Of course, it is preferable to delete the arrows altogether, and use only one, instead of two, terminals to visualize the interconnection, as illustrated in Figure 6.

The one area where signal flows and input/output thinking is perhaps unavoidable is in unilateral systems, such as amplifiers, switches, and other logical devices. These systems possess a signal flow direction, and the behavioral equations do not tell the whole story. For example, the behavioral equation of an amplifier is $(u, y) = (u, Ku)$, where K is the gain, but a complete description involves more than the statement of the joint occurrence of $(u, y = Ku)$. Indeed, even when the behavioral equations are reversible, that is, $K \neq 0$, for such devices we cannot impose the output y and expect the input to follow as $u = y/K$. The functioning of such devices as subsystems requires a proper interconnection architecture. Many systems cannot be backdriven.

An area in physics that deals, by its very nature, with open systems, is thermodynamics. Thermodynamics is considered from an input/output point of view in [14]. However, to formulate the first and second laws of thermodynamics on a suitably general level, inputs and outputs are out of place. There is no reasonable way to partition the external variables, which include work, heat flows, and temperatures, into inputs and outputs. See [15] for an elaboration of this point of view.

In conclusion, input/output thinking and signal-flow diagrams definitely have their place in mathematical modeling and engineering but do not deserve the central place

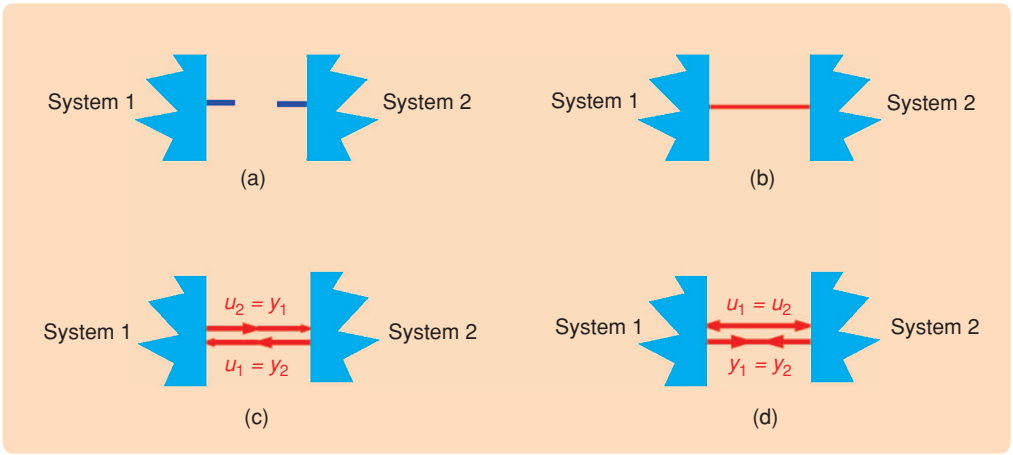


FIGURE 6 Physical system interconnection. Part (a) shows two systems and one terminal of each of the systems. Interconnection by joining the two terminals in (b) leads to sharing terminal variables. In some cases, the terminal variables allow a partition into inputs and outputs with respect to each system. In this case, variable sharing can correspond to output-to-input assignment, as shown in (c). On the other hand, variable sharing may correspond to identifying inputs and identifying outputs, as shown in (d). Viewing system interconnection as output-to-input assignment is not only unnecessary but it is inconvenient when inputs and outputs can be interchanged and is limiting when input/output partitioning is impossible. Variable sharing is the universal and physically justified mechanism for expressing the result of system interconnection through physical terminals.

given to them in systems and control. In addition, variable sharing, not signal transmission is the central and universal idea for interconnection.

TEARING, ZOOMING, AND LINKING

In this section, we outline a formal procedure for obtaining a model by viewing a system, a black box, as an interconnection of subsystems, smaller black boxes. This procedure is useable both in a pedagogical environment and as a blueprint for computer implementations [16], [17]. The problem is to provide a mathematical language for obtaining a model for certain specified variables in an interconnected system from a model of the subsystems, the way in which the systems are interconnected, and the interconnection constraints, keeping in mind Figure 1. The formalism uses the notions of a behavior and of latent variables in an effective way. The basic ingredients are: i) terminals, ii) (parameterized) modules, iii) the interconnection architecture, iv) the module embedding, and v) the manifest variable assignment.

Terminals and Modules

A *terminal* is specified by its *type*. The terminal type may be of a physical nature, such as electrical, mechanical, hydraulic, or thermal type, or logical, such as input or output type. The terminal type implies the nature of the variables that live on this terminal. For example, a voltage and current for a terminal of electrical type, a force and position for a 1D mechanical terminal, a force, position, angle, and torque for a 2D or 3D mechanical terminal, a pressure and mass flow for a hydraulic terminal, a temperature and heat flow for a thermal terminal, an input for a terminal where a variable is imposed on the system, or an output for a terminal where a variable is imposed on the environment. But sometimes a terminal type may involve only a current, or only a voltage, or only a position, depending on the purpose of the model.

A *module* is a dynamical system with a finite number of terminals and a specification of the behavior of the terminal variables. By specifying the *type* of the module, we provide a list of its terminals and their type and therefore a list of the variables that live on the terminals of the module. Usually, the module specification involves, in addition to its type, a set of *parameters*, reflecting the material, geometric, and other properties of the physical device. We assume that, by providing the type of a module and its parameter values, we obtain a specification of the behavior, in the sense of the definition of a dynamical system, of the variables on the terminals of the physical device.

To make concrete what we mean by modules and terminals, we provide a few examples.

Examples of Modules

Consider a 3-ohm resistor, whose module type is *ohmic resistor*. This characterization means that the module has two terminals, both of electrical type, and that the module is parameterized by a nonnegative real number, the value of the resistor in ohms. Since the terminals are electrical, there are two variables, a voltage and a current, counted positive when the current flows into the resistor, on each terminal. In total there are thus four real variables associated with a resistor, namely, (V_1, I_1) and (V_2, I_2). From the fact that we have an ohmic resistor, we know that the relationship among these variables is

$$V_1 - V_2 = RI_1, I_1 + I_2 = 0,$$

where R is the value of the resistor in ohms. Setting $R = 3$ yields the behavioral equations

$$V_1 - V_2 = 3I_1, I_1 + I_2 = 0.$$

These equations completely specify the behavior of an ohmic resistor with parameter value three.

The next example consists of the tanks of black boxes 1 and 3, shown in Figure 3 and discussed in the section "An Illustrative Example." The module type is *tank with two outlets*, which indicates that there are two terminals, both of hydraulic type, and therefore each with an associated pressure and mass flow, counted positive when mass flows into the tank. In total we have four real variables p, f, p', f', as well as the parameters A, B, C, ρ. Equations (1)–(3) describe the behavior of the variables on the terminals of the tank. Note that in a more realistic situation, we would specify the geometry of the tank and the orifices, as well as the material properties of the fluid, and translate this information into the parameters A, B, C, ρ.

The module type of the third example is *proper multivariable transfer function*, with parameters $(\mathtt{m}, \mathtt{p}, G)$, where $\mathtt{m}, \mathtt{p} \in \mathbb{N}$, and G is a $\mathtt{p} \times \mathtt{m}$ matrix of proper real rational functions. This specification indicates a system with $\mathtt{m} + \mathtt{p}$ terminals, the first $\mathtt{m}$ of input type, the last $\mathtt{p}$ of output type, and with behavior described by the controllable input/state/output system

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

with (A, B, C, D) such that $G(\xi) = D + C(I\xi - A)^{-1}B$. In this case there are actually many other ways of translating this module specification into dynamic equations. This class of systems is dealt with extensively in the section "Rational Representations."

The Interconnection Architecture and the Module Embedding

The layout of an interconnected system is visualized as a graph with modules in the vertices and connected terminals as edges (see Figure 7). This layout is formalized by the interconnection architecture and the module

embedding. The *interconnection architecture* or the *interconnection graph* is a graph with leaves. Recall that a *graph* is defined as $\mathcal{G} = (\mathbb{V}, \mathbb{E}, \mathcal{A})$, where $\mathbb{V}$ is a set of *vertices*, $\mathbb{E}$ is a set of *edges*, and $\mathcal{A}$ is the *adjacency map*. The adjacency map $\mathcal{A}$ associates with each edge $e \in \mathbb{E}$ an unordered pair $\mathcal{A}(e) = [v_1, v_2]$ with $v_1, v_2 \in \mathbb{V}$; the edge e is *adjacent* to v_1 and v_2. A graph with leaves [see Figure 7(a)] is a graph in which some special edges, called leaves, are adjacent to only one vertex. Formally, a *graph with leaves* is defined as $\mathcal{G} = (\mathbb{V}, \mathbb{E}, \mathbb{L}, \mathcal{A})$, where $\mathbb{V}$ is a set of *vertices*, $\mathbb{E}$ is a set of *edges*, $\mathbb{L}$ is a set of *leaves*, and $\mathcal{A}$ is the *adjacency map*. The adjacency map $\mathcal{A}$ associates with each edge $e \in \mathbb{E}$ an unordered pair $\mathcal{A}(e) = [v_1, v_2]$ with $v_1, v_2 \in \mathbb{V}$, and with each leaf $\ell \in \mathbb{L}$ an element $\mathcal{A}(\ell) = v \in \mathbb{V}$; e is *adjacent* to v_1 and v_2, while ℓ is adjacent to v. The *degree* of a vertex is the sum of the number of edges and the number of leaves that are adjacent to the vertex. A *self-loop*, that is, an edge with $\mathcal{A}(e) = [v, v]$, contributes 2 to the degree of v.

Modeling an interconnected system requires specifying the laws of the subsystems as well as the interconnection of the subsystems. The concept that formalizes the way in which the subsystems are embedded in the overall system is the *module embedding*, which associates a module with each vertex of the interconnection architecture, as illustrated in Figure 7(b). The degree of the vertex is assumed to be equal to the number of terminals of the associated module. Moreover, the module embedding determines, for every vertex, a one-to-one assignment between the terminals of the module that has been associated with the vertex and the edges and leaves adjacent to the vertex, as illustrated in Figure 7(c). The edges serve to specify how terminals of subsystems are connected, while the leaves allow for unconnected terminals, for example, terminals by which the interconnected system can interact with its environment.

Since each edge is adjacent to two vertices, the module embedding assigns two terminals to each edge. We postulate that this assignment results in two terminals that are of the same type if the terminals are of physical type—both electrical, mechanical, hydraulic, or thermal—or of opposite type—one input, one output—if the terminals are of logical type. In other words, if the edge e is adjacent to vertices v_1 and v_2, then the module embedding must imply that v_1 and v_2 are either of the same physical type, or of opposite logical type. In this way, each vertex is labeled as a module, and each edge and leaf are labeled by a terminal type.

The following examples illustrate interconnection architectures and module assignments.

Example: An RLC Circuit

Consider the electrical circuit of Figure 8. The goal is to model the external port behavior of the circuit. The port variables consist of the difference of the voltages of the external terminals and the current that flows into the circuit along the upper terminal. This circuit has six modules, two ohmic resistors denoted by R_C and R_L, respectively, one capacitor denoted by C, one inductor denoted by L, and two connectors denoted by connector1 and connector2, respectively. The parameter value of modules R_C, R_L, C, and L are denoted by the same symbol as the corresponding module. The parameter value of the modules connector1 and connector2 are both three, meaning that they connect three terminals. All of the terminals of all of the modules are of electrical type, the resistors, capacitor, and inductor each have two terminals, and the connectors each have three terminals. We denote the two terminals of R_C by $R_{C,1}$ and $R_{C,2}$, the 3 terminals of connector1 by connector1_1, connector1_2, and connector1_3, and use a similar notation for the terminals of the other modules.

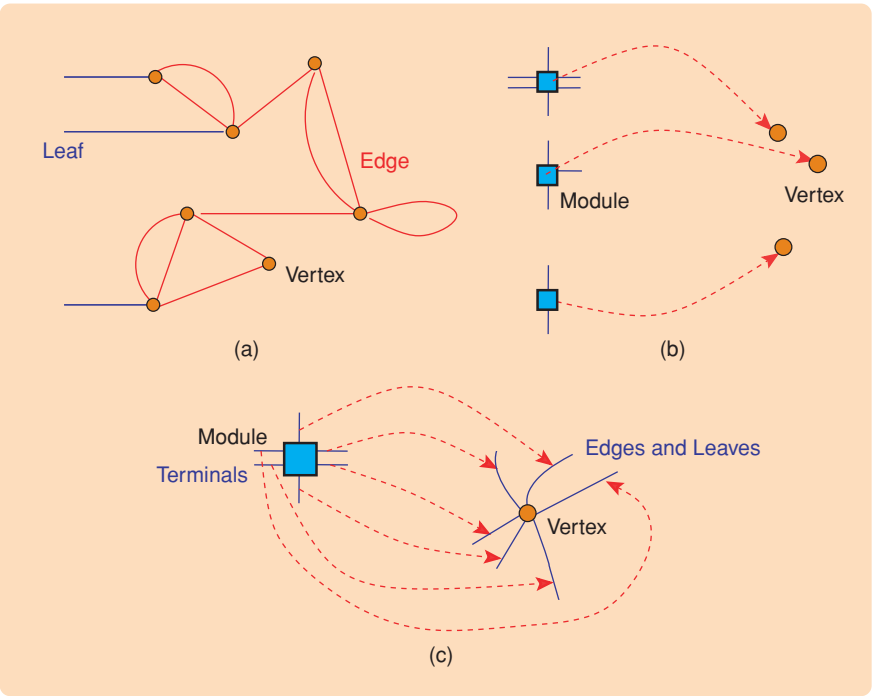


FIGURE 7 Interconnection architecture. The architecture of an interconnected system is formalized in terms of a graph with leaves. Part (a) shows an example of a graph with seven vertices, 11 edges, and three leaves. The module embedding (b) associates with the subsystems one of the vertices of the graph. In (c), each terminal of the subsystem is subsequently assigned to one of the edges or leaves that is adjacent to the corresponding vertex. The interconnection architecture, along with the module embedding, provides a systematic procedure for modeling interconnected systems.

The interconnection architecture, shown in Figure 8(b), has six vertices labeled $1, 2, 3, 4, 5, 6$, six edges labeled c, d, e, f, g, h, and two leaves labeled a, b. The module embedding first requires that we associate a module with each vertex. For the example at hand, this association is given by

$$R_C \mapsto 2,\ R_L \mapsto 5,\ C \mapsto 4,\ L \mapsto 3,$$
$$\text{connector1} \mapsto 1,\ \text{connector2} \mapsto 6.$$

The module embedding also requires that for each vertex, we assign to each edge and leaf adjacent to a vertex, a terminal of the module associated to the vertex. For the RLC example, this assignment is given by

$$\begin{aligned}
\text{vertex 1: } & \text{connector1}_1 \mapsto a,\\
& \text{connector1}_2 \mapsto c,\\
& \text{connector1}_3 \mapsto d,\\
\text{vertex 2: } & R_{C,1} \mapsto c,\ R_{C,1} \mapsto e,\\
\text{vertex 3: } & L_1 \mapsto d,\ L_1 \mapsto f,\\
\text{vertex 4: } & C_1 \mapsto e,\ C_1 \mapsto g,\\
\text{vertex 5: } & R_{L,1} \mapsto f,\ R_{L,1} \mapsto h,\\
\text{vertex 6: } & \text{connector2}_1 \mapsto g,\\
& \text{connector2}_2 \mapsto h,\\
& \text{connector2}_3 \mapsto b.
\end{aligned}$$

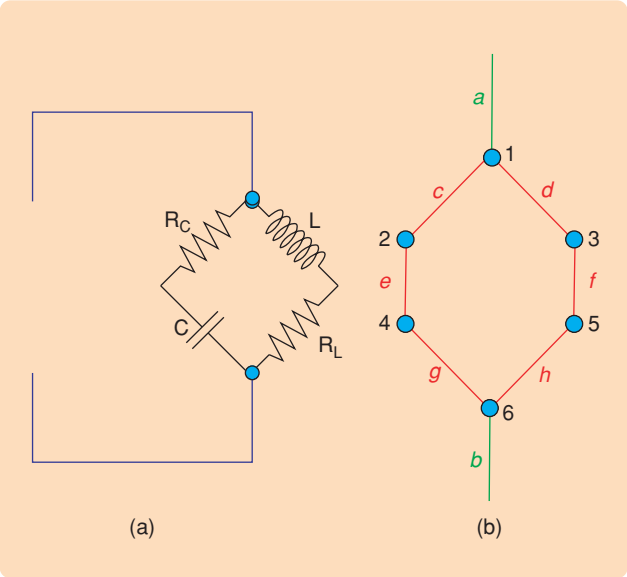
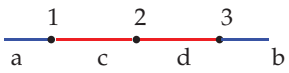


FIGURE 8 Architecture of an RLC circuit. The circuit shown in (a) comprises an interconnection of six subsystems, namely, two resistors, one inductor, one capacitor, and two connectors. The interconnection architecture is shown in (b) as a graph with leaves. The graph has six vertices labeled 1, 2, 3, 4, 5, 6, six edges labeled *c*, *d*, *e*, *f*, *g*, *h*, and two leaves labeled *a*, *b*. The vertices correspond to the subsystems, the edges correspond to the terminals that interconnect the subsystems, and the leaves correspond to the external terminals.

Example: A Hydraulic Example

The next example is the hydraulic system of Figure 3 discussed in the section "An Illustrative Example." This system has three modules, two of type tank with two outlets with parameters A_1, B_1, C_1, ρ_1 and A_3, B_3, C_3, ρ_3, respectively, and one module of type pipe with parameter α. The interconnection architecture is shown below.



The module embedding yields the association

$$\begin{aligned}
&\text{tank}(A_1, B_1, C_1, \rho_1) \mapsto 1,\\
&\text{pipe}(\alpha) \mapsto 2,\\
&\text{tank}(A_3, B_3, C_3, \rho_3) \mapsto 3.
\end{aligned}$$

The assignment of terminals to vertices is

$$\begin{aligned}
\text{vertex 1: } & \text{tank}(A_1, B_1, C_1, \rho_1)_1 \mapsto \text{a},\\
& \text{tank}(A_1, B_1, C_1, \rho_1)_2 \mapsto \text{c},\\
\text{vertex 2: } & \text{pipe}(\alpha)_1 \mapsto \text{c},\ \text{pipe}(\alpha)_2 \mapsto \text{d},\\
\text{vertex 3: } & \text{tank}(A_2, B_2, C_2, \rho_2)_1 \mapsto \text{d},\\
& \text{tank}(A_2, B_2, C_2, \rho_2)_2 \mapsto \text{b}.
\end{aligned}$$

Example: A Feedback System

The third example is the feedback system shown in Figure 9. The interconnection architecture is the graph with vertices A_1, A_2, G_1, G_2, edges $3, 4, 5, 6$, and leaves $1, 2$. The modules consist of two adders, associated with vertices A_1 and A_2, each with two inputs and one output, and two input/output systems, associated with vertices G_1 and G_2. The specification of the module embedding is obvious from Figure 9.

The Interconnection Constraints

The behavioral equations that govern an interconnected system combine module equations with interconnection constraints. We now explain how the interconnection constraints are obtained. The edges of the interconnection architecture specify how the terminals of the modules are linked. A module embedding guarantees that the terminals associated with the same edge are of the same physical type or of opposite logical type.

We postulate that there are universal rules, originating from the physical nature of the interconnections, that specify relations among the variables on the terminals that are linked. For instance, if an edge is electrical type, and hence connects two electrical terminals, the connection rule equates the voltages on the two terminals and equates the

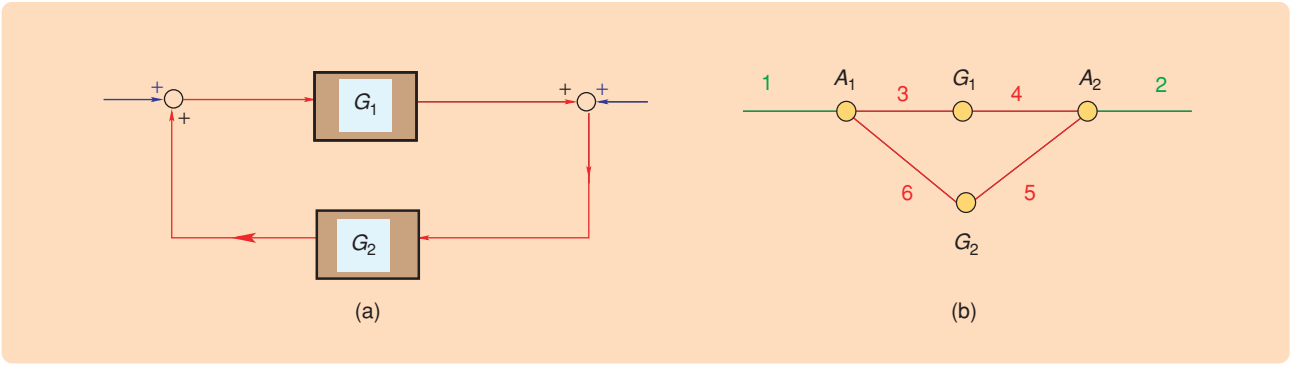


FIGURE 9 Architecture of a feedback system. (a) The feedback system consists of four subsystems, namely, two adders and two input/output systems. This configuration leads to an interconnection architecture represented by (b) the graph with leaves. The graph in (b) has four vertices labeled A_1, A_2, G_1, and G_2, four edges labeled 3,4,5, and 6, and two leaves labeled 1 and 2. Each subsystem is associated with a vertex, while the terminals become edges or leaves of the associated graph. The edges correspond to the internal connections, while the leaves correspond to the external inputs.

sum of the currents on the terminals to zero, where currents are counted positive when they run into a module. If the connected terminals are hydraulic, the connection rule equates the pressures and equates the sum of the mass flows to zero. If the terminals are logical, the connection rule equates the values of the associated input and the associated output.

The behavioral equations of the interconnected system are obtained as follows. For each vertex of the interconnection architecture, we obtain behavioral equations relating the variables that live on the terminals of the module associated with the vertex. These behavioral equations are the *module equations*. For each edge of the interconnection architecture, we obtain behavioral equations relating the variables that live on the terminals and that are linked by the edge. These behavioral equations are the *interconnection equations*, or *interconnection constraints*. Although no interconnection equation results from the leaves, the associated terminal variables nevertheless enter in the module equations.

The module equations and the interconnection equations together specify the behavior of all of the variables on all of the terminals involved. Note that each vertex of the interconnection graph is in the end labeled as a module, while each edge is labeled as a terminal of a specific type. We thus have systems in the vertices and interconnections in the edges in contrast to, for example, conventional electrical circuit theory, which has modules in the edges, and interconnections in the vertices.

The interconnection equations are usually very simple. Typically they equate potential variables and equate the sum of flow variables to zero. We therefore think of interconnection as variable sharing. That interconnections come down to variable sharing is not surprising. In the end, physical interconnection must mean equating volts with volts, kilograms with kilograms, meters with meters, and degrees kelvin with degrees kelvin. Every other interconnection equation must be an abstraction or a transduced version of interconnection constraints involving identically dimensioned physical variables.

For the three examples discussed above we obtain the following specification of the behavior of the terminal variables.

The Module and Interconnection Equations for the RLC Circuit

For the RLC circuit, the module equations involve the currents and voltages on the terminals of the module associated with each of the vertices. The voltage and current of terminal $R_{C,1}$ are denoted by $V_{R_{C,1}}$ and $I_{R_{C,1}}$, respectively, and a similar notation is used for the remaining terminals. The module equations are given by

$$\begin{aligned}
&\text{vertex 1: } V_{\text{connector}_1,1} = V_{\text{connector}_1,2} = V_{\text{connector}_1,3},\\
&\qquad I_{\text{connector}_1,1} + I_{\text{connector}_1,2} + I_{\text{connector}_1,3} = 0;\\
&\text{vertex 2: } V_{R_C,1} - V_{R_C,2} = R_C I_{R_C,1},\ I_{R_C,1} + I_{R_C,2} = 0;\\
&\text{vertex 3: } L\frac{d}{dt} I_{L,1} = V_{L,1} - V_{L,2},\ I_{L,1} + I_{L,2} = 0;\\
&\text{vertex 4: } C\frac{d}{dt}(V_{C,1} - V_{C,2}) = I_{C,1},\ I_{C,1} + I_{C,2} = 0;\\
&\text{vertex 5: } V_{R_L,1} - V_{R_L,2} = R_L I_{R_L,1},\ I_{R_L,1} + I_{R_L,2} = 0;\\
&\text{vertex 6: } V_{\text{connector}_2,1} = V_{\text{connector}_2,2} = V_{\text{connector}_2,3},\\
&\qquad I_{\text{connector}_2,1} + I_{\text{connector}_2,2} + I_{\text{connector}_2,3} = 0.
\end{aligned}$$

The module embedding for the RLC circuit implies that the pairs of terminals

$$\begin{aligned}
&\text{edge } c\text{: } \{R_{C,1}, \text{connector1}_2\}, \quad &&\text{edge } d\text{: } \{L_1, \text{connector1}_3\},\\
&\text{edge } e\text{: } \{R_{C,2}, C_1\}, &&\text{edge } f\text{: } \{L_2, R_{L,1}\},\\
&\text{edge } g\text{: } \{C_2, \text{connector2}_1\}, &&\text{edge } h\text{: } \{R_{L,2}, \text{connector2}_2\}
\end{aligned}$$

share their terminal variables. The interconnection equations, given by

$$\begin{aligned}
&\text{edge c:} \quad V_{R_{C,1}} = V_{\text{connector1}_2},\ I_{R_{C,1}} + I_{\text{connector1,2}} = 0;\\
&\text{edge d:} \quad V_{L_1} = V_{\text{connector1}_3},\ I_{L_1} + I_{\text{connector1}_3} = 0;\\
&\text{edge e:} \quad V_{R_{C,2}} = V_{C_1},\ I_{R_{C,2}} + I_{C_1} = 0;\\
&\text{edge f:} \quad V_{L_2} = V_{R_{C,1}},\ I_{L_2} + I_{R_{L,1}} = 0;\\
&\text{edge g:} \quad V_{C_2} = V_{\text{connector2}_1},\ I_{C_2} + I_{\text{connector2}_1} = 0;\\
&\text{edge h:} \quad V_{R_{L,2}} = V_{\text{connector2}_2},\ I_{R_{L,2}} + I_{\text{connector2}_2} = 0,
\end{aligned}$$

equate the voltages of each of the connected terminals, and equate the sum of the currents to zero.

The module equations together with the interconnection constraints specify the behavior of the terminal variables.

The Module and Interconnection Equations for the Hydraulic Example

For the hydraulic example, the module equations are given by (6)–(8) for vertex 1, (6)–(9) for vertex 2, and (6)–(8) for vertex 3. The interconnection equations are given by (13) for edge c and (14) for edge d.

The Module and Interconnection Equations for the Feedback System

For the feedback system example, we obtain, in the obvious notation, the module equations

$$\begin{aligned}
&\text{vertex } \mathrm{G}_1\colon (u_3, y_4) \in \mathcal{B}_{\mathrm{G}_1}; \quad \text{vertex } \mathrm{G}_2\colon (u_5, y_6) \in \mathcal{B}_{\mathrm{G}_2};\\
&\text{vertex } \mathrm{A}_1\colon y_2 = u_1 + u_6; \quad \text{vertex } \mathrm{A}_2\colon y_4 = u_2 + u_4.
\end{aligned}$$

Here $\mathcal{B}_{\mathrm{G}_1}$ and $\mathcal{B}_{\mathrm{G}_2}$ denote, respectively, the behavior of the input/output systems in the forward loop and the feedback loop of the feedback system. The interconnection equations are given by

$$\begin{aligned}
&\text{edge 3: } y_3 = u_3; \quad \text{edge 4: } y_4 = u_4;\\
&\text{edge 5: } y_5 = u_5; \quad \text{edge 6: } y_6 = u_6.
\end{aligned}$$

The Manifest Variable Assignment

The final step of the modeling procedure consists of the *manifest variable assignment*, a map that assigns the manifest variables as a function of the terminal variables. The terminal variables are henceforth considered as latent variables.

For the RLC circuit the manifest variable assignment consists of the specification

$$\begin{aligned}
V_{\text{externalport}} &= V_{\text{connector}_1,1} - V_{\text{connector}_2,3},\\
I_{\text{externalport}} &= I_{\text{connector1}_1}
\end{aligned}$$

of the external port voltage and port current in terms of the terminal variables. It is easy to deduce from the behavioral equations obtained in the section "The Module and Interconnection Equations for the RLC Circuit" that the equations imply that $I_{\text{connector1}_1} = -I_{\text{connector2,3}}$. In words, the current that flows into the circuit through terminal connector1_1 flows out of the circuit through terminal connector2_3. In many circuit theory applications, modeling aims at obtaining equations of the voltage across a port and the current that flows into a port. We discuss ports and their relation to terminals in the section "Terminals Versus Ports."

For the hydraulic system, the manifest variable assignment is given by (15), while, for the feedback system, the manifest variable assignment consists of

$$u_{\text{external}} = (u_1, u_2),\ y_{\text{external}} = (y_6, y_5).$$

The module equations, combined with the interconnection constraints and the manifest variable assignment, define the full behavior. These equations contain many latent variables—in fact, all of the terminal variables are latent variables—in addition to the manifest variables the model aims at. This model is the end result of the modeling process based on tearing (the interconnection architecture), zooming (leading to the module equations and manifest variable assignment), and linking (leading to the interconnection constraints).

The tearing, zooming, and linking modeling methodology is systematic, modular, adaptable to computer-assisted implementation with the module equations in parametric form and the interconnection equations stored in a database, and hierarchical, since a model of an interconnected systems can be used as a module on a higher level. A model library supporting this methodology is thus reuseable, extendable, modifiable, and flexible. A disadvantage of this methodology is that the model equations involve many variables. This drawback can be alleviated by eliminating variables when possible. The interconnection equations, for example, allow the elimination of many of the variables. In the section "The Elimination Theorem for LTIDS" we explain that, for linear time-invariant systems, a complete elimination of the latent variables can be carried out using computer algebra algorithms.

The philosophy of tearing, zooming, and linking is to keep the interconnections highly standardized and simple, and to deal with complex features of a model by means of modules. For instance, in the circuit example, a multiterminal connector is viewed as a module, rather than as a connection. In mechanical systems, joints, hooks, and hinges are viewed as modules, rather than as connections. As a caveat, we emphasize that not all interconnections fit this framework. In particular, distributed interconnections, such as mechanical systems interconnected by sharing a flexible surface, or heat conduction along a surface, do not fit the framework described, since we assume a finite set of terminals. Also, terminals do not immediately capture interconnections along virtual terminals, for example, action at a distance, such as the attraction of masses due to gravity or electrical charge. Finally, interactions such as rolling, sliding, and bouncing also require a different framework. The variable-sharing approach of tearing,

zooming, and linking formalizes the modeling practice followed in computer-assisted modeling packages such as Spice [18] and Modelica [19], in contrast to Matlab's output-to-input assignment-based Simulink [20].

BOND-GRAPH MODELING

The modeling philosophy of tearing, zooming, and linking has a great deal of affinity with *bond graphs* [21]. The central notion in bond graphs is a *bond*, which is interpreted as a connection between a pair of subsystems. The idea that energy is exchanged between subsystems along a connection is a central feature of bond-graph modeling. Associated with a bond, there is an *effort* variable, denoted by e, and a *flow* variable, denoted by f. The product ef, or, more generally, the inner product $\langle e, f \rangle$ in the case of multibonds, is postulated to be the rate of physical energy that flows from one subsystem to another subsystem along the bond. A half arrow (harpoon) is used to indicate the direction of positive energy flow. Examples are electrical bonds with voltage as effort and current as flow, mechanical bonds with force and torque as effort and velocity and angular velocity as flow, and hydraulic bonds with pressure as effort and mass flow as flow.

A subsystem imposes a relation among the effort and flow variables associated with it. Interconnections are formalized by means of *junctions*. A parallel junction of $\mathbf{n}$ bonds imposes the equations

$$e_1 = e_2 = \cdots = e_{\mathbf{n}}, \quad f_1 + f_2 + \cdots + f_{\mathbf{n}} = 0,$$

as, for example, the relations imposed by a connection of $\mathbf{n}$ wires in electrical circuits. A series junction of $\mathbf{n}$ bonds imposes

$$f_1 = f_2 = \cdots = f_{\mathbf{n}}, \quad e_1 + e_2 + \cdots + e_{\mathbf{n}} = 0.$$

Using subsystems and junctions as building blocks, the bond graph modeling methodology applies to many physical domains and has been successfully incorporated in computer-aided modeling packages.

As a modeling methodology, bond graphs are much more realistic and general than the output-to-input assignment of conventional system theory. In its basic philosophy, the bond-graph approach is similar to the variable-sharing approach. A bond is like a terminal, while junctions play the role of interconnection constraints. Tearing, zooming, and linking is more general than bond graphs, since it is not based on power and energy considerations, although the interconnection constraints automatically recover conservation of energy whenever this property is relevant.

Bond graphs postulate that the variables on the terminals that are shared by interconnected systems have certain properties. We now examine two questions that come up in the axiomatics underlying bond graphs.

1) To what extent are the efforts and flows that characterize the terminal variables in bond graphs universal?
2) Do transmission and conservation of energy play the central role in characterizing interconnections that bond-graph thinking attributes to it?

Efforts and flows are related to across and through variables as well as to intensive and extensive quantities in thermodynamics. The idea that terminal variables can be divided into efforts and flows appears to be a deep—but unfortunately largely unexplored—physical principle, corroborated by many examples. As we have seen, electrical, mechanical, hydraulic, and heat flow terminals all involve effort variables and flow variables. Moreover, typical interconnection constraints equate efforts and equate the sum of flows to zero.

What is the physical or mathematical background of these efforts and flows? One possible explanation is as follows. Consider the interconnection of two terminals of the same type. Denote the variables of the first terminal by $w_1 \in \mathbb{R}^{\mathbf{n}}$ and of the second terminal by $w_2 \in \mathbb{R}^{\mathbf{n}}$. Let $\mathcal{C}$ be the subspace of $\mathbb{R}^{\mathbf{n}} \times \mathbb{R}^{\mathbf{n}}$ that expresses the interconnection constraint, assumed linear. After interconnection, the variables w_1, w_2 have to satisfy $(w_1, w_2) \in \mathcal{C}$. It is natural to assume that interconnection imposes symmetric constraints on the variables w_1 and w_2, that is,

$$(w_1, w_2) \in \mathcal{C}$$

if and only if

$$(w_2, w_1) \in \mathcal{C}.$$

Assume also that $\texttt{dimension}(\mathcal{C}) = \mathbf{n}$ and that interconnection constraint leaves the individual variables w_1 and w_2 free, that is, for all $w_1 \in \mathbb{R}^{\mathbf{n}}$, there exists $w_2 \in \mathbb{R}^{\mathbf{n}}$ such that $(w_1, w_2) \in \mathcal{C}$, and, for all $w_2 \in \mathbb{R}^{\mathbf{n}}$, there exists $w_1 \in \mathbb{R}^{\mathbf{n}}$ such that $(w_1, w_2) \in \mathcal{C}$. These conditions imply that there exists a basis in $\mathbb{R}^{\mathbf{n}}$ such that, in this basis, $w_1 = (e_1, f_1)$, $w_2 = (e_2, f_2)$, and the interconnection constraint $(w_1, w_2) \in \mathcal{C}$ becomes

$$e_1 = e_2, \quad f_1 + f_2 = 0.$$

Indeed, since $\texttt{dimension}(\mathcal{C}) = \mathbf{n}$, and since, for all $w_1 \in \mathbb{R}^{\mathbf{n}}$, there exists $w_2 \in \mathbb{R}^{\mathbf{n}}$ such that $(w_1, w_2) \in \mathcal{C}$, the interconnection constraint $(w_1, w_2) \in \mathcal{C}$ is of the form

$$w_2 = Cw_1$$

for a matrix $C \in \mathbb{R}^{\mathbf{n} \times \mathbf{n}}$. Symmetry of the interconnection constraints implies that $C^2 = I_{\mathbf{n} \times \mathbf{n}}$, that is, C is an involution. In a suitable basis, therefore,

$$C = \begin{bmatrix} I_{\mathbf{n}_+ \times \mathbf{n}_+} & 0 \\ 0 & -I_{\mathbf{n}_- \times \mathbf{n}_-} \end{bmatrix}.$$

In other words, the existence of efforts and flows follows from certain general properties of the interconnection constraint, in particular, from symmetry of the interconnection constraint. More structure is required, however, to deduce that the dimension of the space of effort variables is equal to the dimension of the space of flow variables, as assumed in bond-graph modeling.

It is more difficult to understand the physical origin of the assumption that the inner product of effort and flow that enter in the interconnection constraints equals power, the rate of energy transmitted along the terminal. I give four reasons for my skepticism (see "Bond Graphs" for a summary).

The assumption that the product of effort and flow is power is sometimes not natural. For a thermal terminal, the natural interconnection constraints are

$$T_1 = T_2, \quad Q_1 + Q_2 = 0,$$

where T_1 and T_2 denote temperatures, and Q_1 and Q_2 denote heat flows. Although temperature is a valid effort variable and heat flow is a valid flow variable, Q is the rate of energy transmitted, and the product of effort and flow is not power. Consequently, bond-graph modeling requires the equivalent interconnection constraints

$$T_1 = T_2, \quad \frac{Q_1}{T_1} + \frac{Q_2}{T_2} = 0,$$

with the second equation interpreted as entropy flow. Requiring that the product of effort and flow be power seems artificial in this example.

The Product of Effort and Flow

The assumption in bond-graph modeling that the product of effort and flow is power is sometimes not true. For example, for a 1D mechanical terminal the physical interconnection constraints are

$$p_1 = p_2, \quad F_1 + F_2 = 0,$$

where the effort variables p_1 and p_2 are the positions of the connected terminals, and the flow variables F_1 and F_2 are the

Bond Graphs

The tearing, zooming, and linking methodology for modeling interconnected systems advocated and developed in this article has many things in common with bond graphs. Introduced by Paynter [S4] in the 1960s, bond graphs are popular as a methodology for modeling interconnected physical systems, especially in mechanical engineering (see [S5] and [S6] and for a recent exposition, see [22]). For modeling physical systems, bond-graph modeling is a superior alternative to signal-flow diagrams and input/output-based modeling procedures.

Bond graphs view each system interconnection in terms of power and energy. The variables associated with terminals are assumed to consist of an *effort* and a *flow*, where the (inner) product of effort and flow is *power*. Connections are formalized by *junctions*. Using a combination of junctions and component subsystems, complex physical systems can be modeled in a systematic way. The power interpretation automatically takes care of conservation of energy. The philosophy underlying bond graphs is, as stated in [22],

Power is the universal currency of physical systems.

The idea that terminal variables come in pairs, an effort and a flow, with efforts preserved at each interconnection and the sum of flows equated to zero at each interconnection, is appealing and deep. But, in addition to weak mathematical underpinnings and unconventional graph notation with half arrows, bond graphs have some shortcomings as a modeling philosophy, as explained in the section "Bond-Graph Modeling." The main points discussed in that section are the following:

1) The requirement that the product of effort and flow must be power is sometimes not natural, for example, in thermal interconnections.
2) In connecting terminals of mechanical systems, bond-graph modeling equates velocities, and sets the sum of the forces equal to zero. In reality one ought to equate positions, not velocities. Equating velocities instead of positions leads to incomplete models.
3) Interconnections are made by means of terminals, while energy is transferred through ports. Ports involve many terminals simultaneously. The interconnection of two electrical wires involves equating two terminal potentials and putting the sum of two terminal currents to zero. The product of effort, namely, the electrical potential, and flow, namely, the electrical current, for an electrical connection has the dimension of power, but it is not power. Power involves potential differences, while the interconnection constraints involves the terminal potentials themselves. It is not possible to interpret these interconnection constraints as equating the power on both sides of the interconnection point.
4) In many interconnections, it is unnecessary to have to worry about conservation of energy.

REFERENCES

[S4] H. Paynter, *Analysis and Design of Engineering Systems*. Cambridge, MA: MIT Press, 1961.

[S5] D.C. Karnopp, D.L. Margolis, and R.C. Rosenberg, *System Dynamics: A Unified Approach*. New York: Wiley-Interscience, 1990.

[S6] F.E. Cellier, *Continuous System Modeling*. New York: Springer-Verlag, 1991.

forces acting on the connected terminals. In this case, the product of effort and flow is not power. Guided by the requirement that the product of effort and flow must be power, bond-graph modeling insists on the interconnection constraints

$$v_1 = v_2, \quad F_1 + F_2 = 0,$$

where v_1 and v_2 are the velocities of the connected terminals. Equating positions of course implies equating velocities, but the converse is not valid. So the insistence on having the product of effort and flow equal power leads to incomplete and thus incorrect interconnection equations. To illustrate this difficulty, consider the problem of obtaining the behavior of the position of the mass of the example shown in Figure 10. This example is taken from [22, Figure 2]. The parameters m and c denote, respectively, the mass and the compliance of the spring. The bond-graph equations (see [22, (3)–(6)]) are

$$\begin{aligned}
&\text{Mass:} && f_1 = \frac{p}{m}, && \frac{d}{dt}p = e_1,\\
&\text{Spring:} && e_2 = \frac{q}{c}, && \frac{d}{dt}q = f_2,\\
&\text{Interconnection:} && e_1 = e_2, && f_1 = -f_2.
\end{aligned}$$

Since we are interested in the position w of the mass, we add the manifest variable assignment

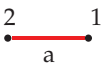
$$\frac{d}{dt}w = f_1.$$

To obtain the behavioral equation for w, we eliminate the latent variables f_1, f_2, e_1, e_2, p, q to obtain

$$mc\frac{d^3}{dt^3}w + \frac{d}{dt}w = 0 \tag{21}$$

as the equation of motion for w obtained by bond-graph modeling.

Tearing, zooming, and linking applied to this example yields the interconnection architecture



with the mass in vertex 1 and the spring in vertex 2, interconnected by edge a. With F_1 the force acting on the mass, E_1 the position of the mass, F_2 the force acting on the spring, and E_2 the position of the endpoint of the spring, we obtain the equations

$$\begin{aligned}
&\text{Mass:} && F_1 = m\frac{d^2}{dt^2}E_1,\\
&\text{Spring:} && E_2 = \frac{F_2}{c},\\
&\text{Interconnection:} && E_1 = E_2,\ F_1 = -F_2,\\
&\text{Manifest variable assignment:} && w = E_1.
\end{aligned}$$

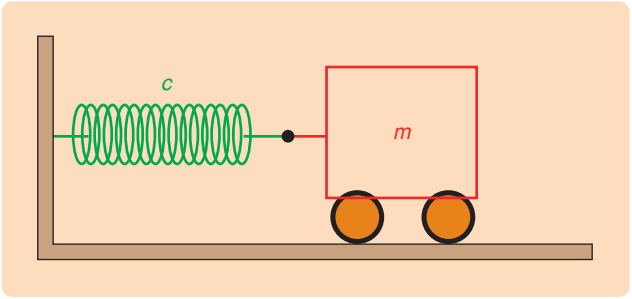


FIGURE 10 Bond-graph modeling. This mechanical system, which is discussed in [21, Figure 2], consists of the interconnection of a mass m and a spring with compliance c. In bond-graph modeling, as well as in other methodologies that make conservation of energy the central mechanism of interconnection, the connection constraints equate the velocities (flows) and forces (efforts) of the terminals at the interconnection point. Since in reality the interconnection constraints equate positions rather than velocities, this procedure may lead to an incomplete model.

After eliminating F_1, E_1, F_2, E_2, we obtain

$$mc\frac{d^2}{dt^2}w + w = 0 \tag{22}$$

as the equation for w. The behavior of (21) consists of all trajectories $w : \mathbb{R} \to \mathbb{R}$ of the form

$$w(t) = A_0 + A_1(\cos t/\sqrt{cm}) + A_2\cos(t/\sqrt{cm}),$$

with A_0, A_1, A_2 ranging over $\mathbb{R}$, whereas the behavior of (22) consists of all trajectories $w : \mathbb{R} \to \mathbb{R}$ of the form

$$w(t) = A_1\cos(t/\sqrt{cm}) + A_2(\cos t/\sqrt{cm}),$$

with A_1, A_2 ranging over $\mathbb{R}$.

The difference between the behavior for the position of the mass w obtained by bond-graph modeling and by tearing, zooming, and linking consists of the constant A_0. This difference is due to the fact that equating only velocities allows the possibility of a spurious constant difference between the positions of the two interconnected terminals. Equating only velocities rather than positions is simply not what the physics demands. Mechanical connections involve more than conservation of energy.

Often, the reply to this argument is that this difference does not matter, since in simulations the spurious constant disappears by setting the correct initial conditions. This reasoning is not convincing, since models must predict what motions are possible. Bond-graph modeling leads to too many elements in the behavior and therefore to incomplete models. In addition, there are things that matter for models other than simulations with specific initial conditions. For instance, including damping in (22) yields a system that is asymptotically stable, whereas including

damping in (21) does not yield asymptotic stability, since the spurious constant still appears when damping is present. Requiring that the product of the interconnection variables must be power can, as shown by this example, lead to an incomplete model. This same problem is also present in port-Hamiltonian systems [23]. For example, the interconnection of two masses using energy conservation yields [23, (2.112)], which has the same spurious constant in the behavior of the position of the connected masses.

Terminals Versus Ports

In electrical circuits, energy is not transmitted along terminals but rather through ports. On the other hand, in modeling and interconnection, terminals matter, not ports. In this subsection, we explain the difference between terminals and ports. This difference is a subtle matter, but it is essential for understanding the limitations of bond graphs relative to the tearing, zooming, and linking methodology.

Consider the electrical circuit illustrated in Figure 11. Associated with each terminal are two real numbers, a potential and a current, counted positive when the current flows into the circuit. Assume that there are $\mathtt{n}$ terminals, each of which is a wire to which a terminal of another circuit can be connected. Denote the $\mathtt{n}$-vector of terminal potentials by V, and the $\mathtt{n}$-vector of terminal currents by I. The $\mathtt{n}$-terminal circuit defines, through its terminal behavior, a dynamical system $(\mathbb{R}, \mathbb{R}^{\mathtt{n}} \times \mathbb{R}^{\mathtt{n}}, \mathcal{B})$.The behavior $\mathcal{B} \subseteq (\mathbb{R}^{\mathtt{n}} \times \mathbb{R}^{\mathtt{n}})^{\mathbb{R}}$, to qualify as the terminal behavior of an electrical circuit that stores no net electrical charge, must satisfy *Kirchhoff's laws*, expressed as

$$(V, I) \in \mathcal{B} \quad \text{and} \quad \alpha : \mathbb{R} \to \mathbb{R}$$

imply

$$(V + \alpha \mathtt{e}, I) \in \mathcal{B} \quad \text{and} \quad \mathtt{e}^{\top} I = 0, \tag{23}$$

where $\mathtt{e} = [1\ 1 \cdots 1\ 1]^{\top}$ is the $\mathtt{n}$-vector of 1's. The condition $(V + \alpha \mathtt{e}, I) \in \mathcal{B}$ states that the behavioral equations of the electrical circuit involve only the potential differences $V_{\mathtt{k}} - V_{\ell}$ of the terminal potentials. The potentials are not

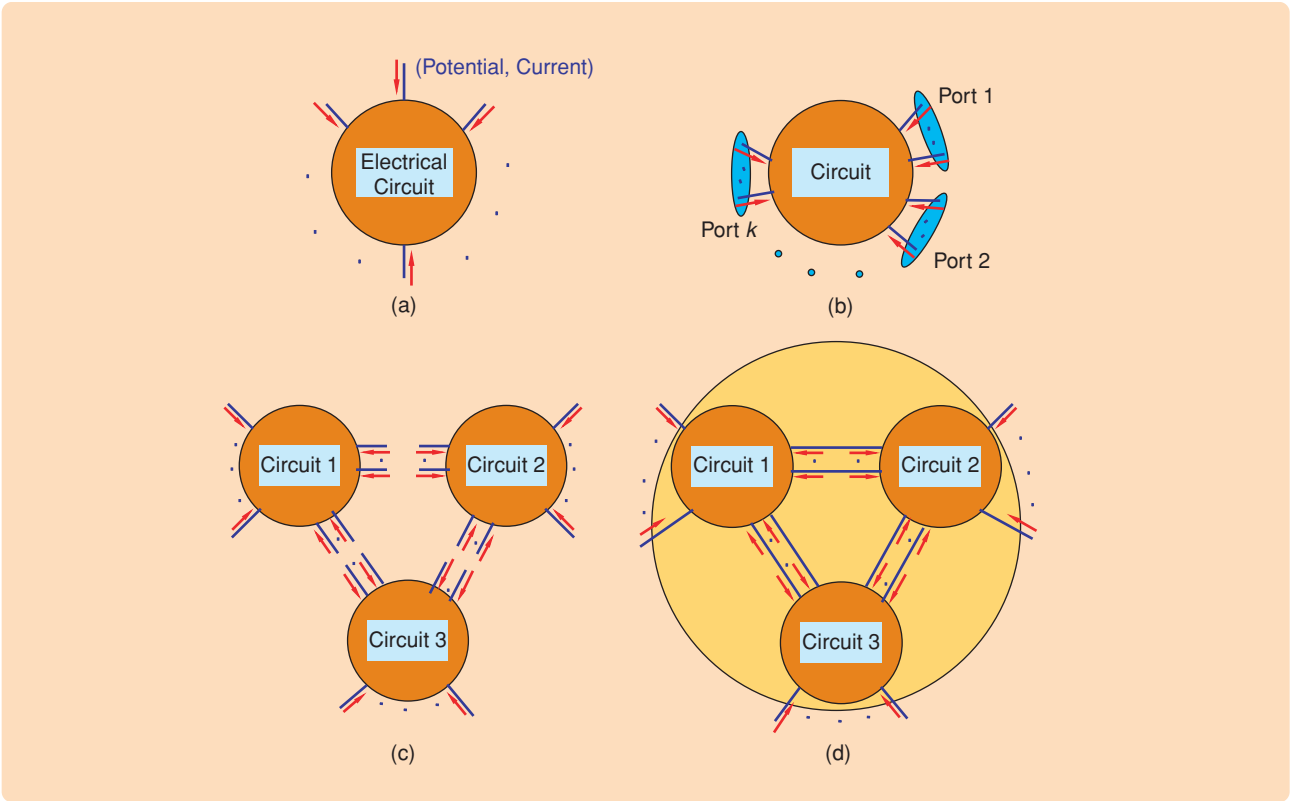


FIGURE 11 Terminals versus ports in electrical circuits. Part (a) shows an electrical circuit that can interact with its environment through terminals in the form of wires. Each wire is a terminal with a potential and a current as associated terminal variables. The behavior of the circuit consists of the pairs of terminal potential/current vectors that meet the dynamical laws of the circuit. A subset of terminals such that potential/current behavior of the circuit implies that the sum of the currents into the circuit along the terminals of the subset is zero and such that the behavior is invariant under the addition of a constant to the potentials of the terminals of the subset, is a *port*. A circuit may possess several ports, as illustrated in (b). Energy transfer in an electrical circuit occurs through ports. It is not possible in general to state what the energy transferred from the environment to a circuit is along a subset of terminals that do not form a port. Circuits are interconnected by means of terminals, as shown in (c) and (d). When the interconnected terminals of Circuit 1 and Circuit 2 do not form a port, it is not possible in general to state what the energy is that flows from Circuit 1 to Circuit 2.

measurable physical quantities; only their differences are. The condition $\mathbf{e}^\top I = 0$ states that there is no net accumulation of charge in the circuit. The *port conditions* (23) imply

$$V^\top I = (V + \alpha \mathbf{e})^\top I.$$

Hence $V^\top I$ is independent of α and is indeed a physical quantity, namely the power, the rate of energy that flow into the circuit.

More generally, a subset of $\mathbf{n}'$ terminals of an electrical circuit is a *port* if

$$\left(\begin{bmatrix} V' \\ V'' \end{bmatrix}, \begin{bmatrix} I' \\ I'' \end{bmatrix} \right) \in \mathcal{B} \quad \text{and} \quad \alpha : \mathbb{R} \to \mathbb{R}$$

imply

$$\left(\begin{bmatrix} V' + \alpha \mathbf{e}' \\ V'' \end{bmatrix}, \begin{bmatrix} I' \\ I'' \end{bmatrix} \right) \in \mathcal{B} \quad \text{and} \quad \mathbf{e}'^\top I' = 0,$$

where $\mathbf{e}' = [1\ 1 \cdots 1\ 1]^\top$ is the $\mathbf{n}'$-vector of 1's, V' is the $\mathbf{n}'$-vector of potentials, and I' is the $\mathbf{n}'$-vector of currents corresponding to the subset of terminals. This way, the complete set of terminals of a circuit can be partitioned into $\mathbf{k}$ disjoint subsets, each of which forms a port. Such a circuit is a $\mathbf{k}$-*port*, as illustrated in Figure 11. $V'^\top I'$ is the rate of energy that flows into the circuit along the port. The total energy flow into the circuit is the sum of the energy flows along the ports. The basic electrical circuit elements consist of two-terminal one ports, such as resistors, capacitors, and inductors, as well as four-terminal two ports, such as transformers and gyrators. An $\mathbf{n}$-terminal connector

$$V_1 = V_2 = \cdots = V_{\mathbf{n}}, \quad I_1 + I_2 + \cdots + I_{\mathbf{n}} = 0$$

is a one port. By interconnecting these building blocks, we can obtain circuits with any number of ports and with any number of terminals for each port. For example, the symbols Y or Δ are often used to denote three-terminal one ports.

The interconnection of electrical circuits that satisfy the port conditions (23), using the interconnection equations

$$V_1 = V_1, \quad I_1 + I_2 = 0$$

terminal by terminal, again satisfies the port conditions. The port conditions (23) hold not only if we consider only the external terminals and view interconnection variables as latent variables but also if we keep interconnected terminals and the interconnection variables in the picture, as illustrated in Figure 11 for the case $\mathbf{k} = 3$. After interconnection, circuits 1, 2, and 3 remain ports, while the interconnection of circuits 1, 2, and 3 is also a port when we consider only the external terminals. However, there is no reason for a subset of the external terminals, say, the external terminals that are terminals of circuit 1, to form a port, or for the terminals of circuit 1 that are connected to terminals of circuit 2 to form a port. We therefore cannot speak of the energy transferred from the environment to circuit 1, or of the energy transferred from circuit 1 to circuit 2 in the interconnected circuit.

Terminals take care of interconnections, whereas ports take care of energy transfer. If a set of terminals forms a port, then the rate of energy transferred into the circuit along this port is given by $V'^\top I'$. If a set of terminals does not form a port, then $V'^\top I'$ has the dimension of power (its dimension is watt = volts × amps) but $V'^\top I'$ does not necessarily represent the rate of energy flow into the circuit along the set of terminals. Conservation of energy in the interconnected circuit shown in Figure 11 is a consequence of the equality

$$V_{\mathrm{e}}^\top I_{\mathrm{e}} = V_1^\top I_1 + V_2^\top I_2 + V_3^\top I_3,$$

where $V_{\mathrm{e}}, I_{\mathrm{e}}$ denote the voltage and current vector of the external terminals, V_1, I_1 of circuit 1, V_2, I_2 circuit 2, and V_3, I_3 of circuit 3. This equality follows from the interconnection constraint. But we cannot express conservation of energy by stating that the energy that flows from circuit 1 into the interconnection points with circuit 2 is equal to the energy that flows from the interconnection points into circuit 2. Although the interconnection equations imply conservation of energy, it is not possible to interpret an interconnection constraint as equating the flow of energy from one side of the interconnection wire to the other side

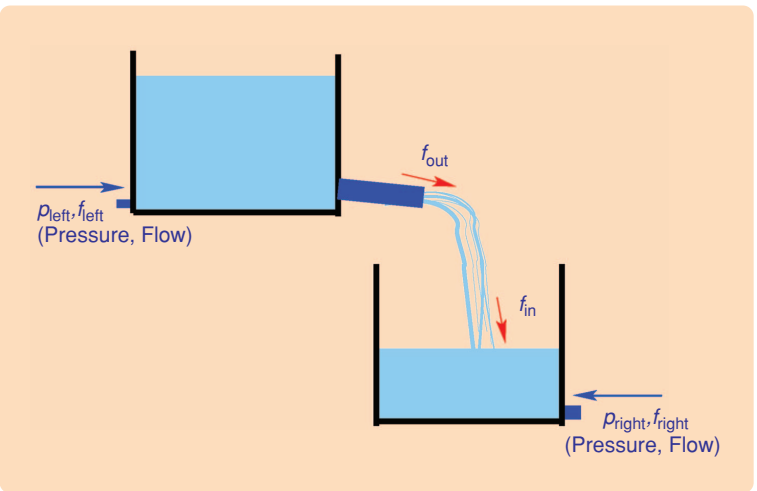


FIGURE 12 Conservation of energy in interconnections. In some situations, such as this hydraulic system, conservation of energy is irrelevant to modeling the interconnections. Interconnection of the two tanks requires that the flow out of the upper tank equal the flow into the lower tank; that is, conservation of mass. There is no need to introduce additional interconnection variables that allow interpretation of the interconnection constraints as conservation of energy.

of the interconnection wire, contrary to the basic bond-graph philosophy.

A final reason for the restrictiveness of bond-graph modeling concerns physical interconnections where conservation of energy is not an issue. For example, for the system in Figure 12, it is unnecessary to include conservation of energy as a consequence of the interconnection constraints. In this example, conservation of mass suffices.

In conclusion, modeling interconnected systems by bond graphs is well motivated and has a much more compelling physical foundation than output-to-input assignment of classical system theory. The premise that physical interconnections involve paired efforts and flows appears deep. However, the physical origin of the effort and flow variables, while clear in many examples, seems elusive as a general principle. Finally, there are reasons to doubt the universal interpretation of the inner product of effort and flow in an interconnection as power.

STATE REPRESENTATIONS

The state of a dynamical system is a central concept in areas such as control, physics, automata, discrete-event systems, and algorithms. The notion of state formalizes the memory of a system and is a clear idea, much more so than elusive concepts such as cause and effect. In behavioral theory the state is viewed as a special type of latent variable. But, as we have argued throughout this article, first principle models lead to equations with many latent variables other than the state variables. While manifest variables generalize traditional inputs and outputs, latent variables generalize traditional states. This generalization is already required in elementary physical examples. There is no reason to give the state the central role in models for dynamical systems that it has been given. The state is a special, and very useful, latent variable that is constructed such that the past and future trajectories become conditionally independent given the present state, with independence understood in a deterministic, set-theoretic, sense.

The notion of state is formalized as follows. The latent-variable dynamical system $\Sigma_{\text{full}} = (\mathbb{T}, \mathbb{W}, \mathbb{X}, \mathcal{B}_{\text{full}})$ with $\mathcal{B}_{\text{full}} \subseteq (\mathbb{W} \times \mathbb{X})^{\mathbb{T}}$ is a *state system* if

$$(w_1, x_1) \in \mathcal{B}_{\text{full}}, (w_2, x_2) \in \mathcal{B}_{\text{full}}, t_0 \in \mathbb{T}, \text{ and } x_1(t_0) = x_2(t_0)$$

imply

$$(w_1, x_1) \wedge_{t_0} (w_2, x_2) \in \mathcal{B}_{\text{full}}, \tag{24}$$

where $\wedge_{t_0}$ denotes *concatenation*, defined by

$$f_1 \wedge_{t_0} f_2(t) := \begin{cases} f_1(t) & \text{if } t < t_0, \\ f_2(t) & \text{if } t \geq t_0. \end{cases}$$

The state definition implies that, given the state at a certain time, the concatenation of any legal trajectory leading to that state with any legal trajectory emanating from that state is again a legal trajectory. Concatenability is what is meant by conditional independence of the past and the future, given the present state.

It is easy to see that a discrete-time system with full behavior described by the behavioral difference equation

$$f(w(t), x(t), x(t+1)) = 0 \quad \text{for all} \quad t \in \mathbb{T},$$

is a state system. Similarly, for continuous-time systems, the system described by the differential equation

$$f\left(w(t), x(t), \frac{d}{dt}x(t)\right) = 0 \quad \text{for all} \quad t \in \mathbb{T},$$

also defines a state system, provided a suitable notion of solution is used.

For continuous-time systems described by differential equations, it is sometimes convenient to assume that the behavior consists of smooth, for example, infinitely differentiable, solutions of the differential equation. However, this smoothness assumption may be in conflict with the concatenability requirement for the state property, since the concatenation of two infinitely differentiable maps is usually not infinitely differentiable. In this case, condition (24) is usually relaxed to

$$(w_1, x_1) \wedge_{t_0} (w_2, x_2) \in \bar{\mathcal{B}}_{\text{full}},$$

where $\bar{\mathcal{B}}_{\text{full}}$ denotes the closure of $\mathcal{B}_{\text{full}}$ in a suitable topology, for example, uniform continuity on finite intervals.

The problem of constructing a state realization for a behavior is studied in [2] for general systems. Concrete algorithms that pass from a behavior to a state representation are derived for systems described by linear time-invariant differential or difference in [1] and [24]–[27]. The ideas underlying these algorithms are discussed in "The Initial Value Problem."

CONTROLLABILITY AS A SYSTEM PROPERTY

The development of modern control theory received a sharp impulse through the introduction by R.E. Kalman of the notion of controllability [28]. In "A Brief History of Controllability," we sketch the evolution of this notion. In its usual setting, controllability refers to the ability to transfer the state of a controlled system between any two points (see Figure 13). Controllability has strong appeal, both from a control engineering and mathematical point of view. State controllability asks a compelling question: *if a system has drifted into an undesirable state, can it be steered back to another desirable one? Can we reach any state from any other state?*

Involving the state in the definition of controllability is a drawback for at least the following related reasons:

1) Since first principles models are seldom in state form, this definition is not applicable to a set of dynamical equations describing a specific behavior. In other words, state controllability is not a system property but a property of a state representation.

2) The state is a construct. If a model is not in state form, we can apply a state construction procedure that brings the model in state form. Assume now that the resulting state model is not state controllable. *What is this lack of controllability then due to?* It may be that the control has indeed not enough influence on the system. But lack of state controllability could also be due to a bad choice of the state. Lack of state controllability therefore does not really tell us that the control has inadequate influence on the system dynamics.
3) It is difficult to see how this definition can be generalized to situations in which the state property is not well developed, for example, to systems in which the set of independent variables involves both space and time, as in PDEs.

The behavioral theory of systems provides a notion of controllability that circumvents these drawbacks. For ease of exposition, controllability is defined here only for time-invariant systems with time axis $\mathbb{T} = \mathbb{R}$ or $\mathbb{Z}$.

The time-invariant dynamical system $\Sigma = (\mathbb{T}, \mathbb{W}, \mathcal{B})$, with $\mathbb{T} = \mathbb{R}$ or $\mathbb{Z}$, is *controllable* if, for all $w_1, w_2 \in \mathcal{B}$, there exist $w : \mathbb{T} \to \mathbb{W}$ and $t' \in \mathbb{T}$ such that i) $w \in \mathcal{B}$, ii) $w(t) = w_1(t)$ for all $t < 0$, and iii) $w(t + t') = w_2(t)$ for all $t \geq 0$. In words, we start with two trajectories $w_1, w_2 \in \mathcal{B}$, in the behavior, and we want to transfer from the undesired trajectory w_1 to the desired trajectory w_2. The trajectory w executes this transfer, in t' units of time, and yields a trajectory that has the past of w_1 as its past, and the future of w_2 as its future. If this maneuver is possible for all $w_1, w_2 \in \mathcal{B}$, using a legal trajectory $w \in \mathcal{B}$, then the system is controllable. This definition is illustrated in Figure 13. It is easy to see that state controllability is a special case of behavioral controllability, by taking either $w = x$ or $w = (u, x)$.

The definition of controllability is readily generalized to stabilizability. The dynamical system $\Sigma = (\mathbb{T}, \mathbb{W}, \mathcal{B})$, with $\mathbb{W}$ a subset of a normed vector space, is *stabilizable* if, for all $w \in \mathcal{B}$, there exists $w' \in \mathcal{B}$ such that $w(t) = w'(t)$ for all $t < 0$ and $w'(t) \to 0$ as $t \to \infty$. In stabilizability, we want to transfer from an undesired trajectory w to a desired trajectory w' that goes to zero. If this maneuver is possible for all $w \in \mathcal{B}$, using a legal trajectory $w' \in \mathcal{B}$, then the system is stabilizable.

Consider the system $\Sigma = (\mathbb{T}, \mathbb{W}_1 \times \mathbb{W}_2, \mathcal{B})$. Then w_2 is *observable from* w_1 in Σ if

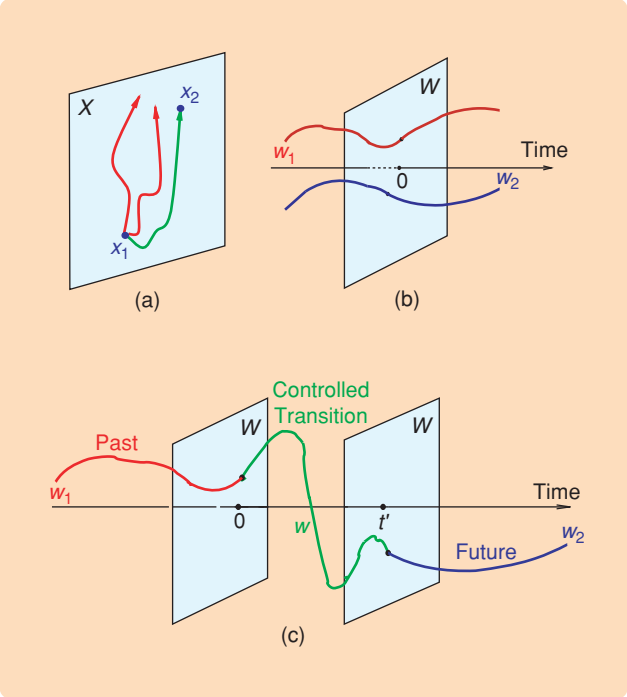


FIGURE 13 Behavioral controllability. The classical notion of state controllability requires, for each initial and terminal state, the existence of an input that takes the state of a system from the initial to the terminal state. This notion is illustrated in (a). Behavioral controllability for a dynamical system is illustrated in (b) and (c). Part (b) shows two time trajectories in the behavior. The system is controllable if there is a trajectory in the behavior that has the past of the first trajectory as its past and the future of the second trajectory as its sometime future, as shown in (c). Combined with the definition of a dynamical system as a behavior, controllability becomes a genuine property of a dynamical system rather than a property of just a state representation.

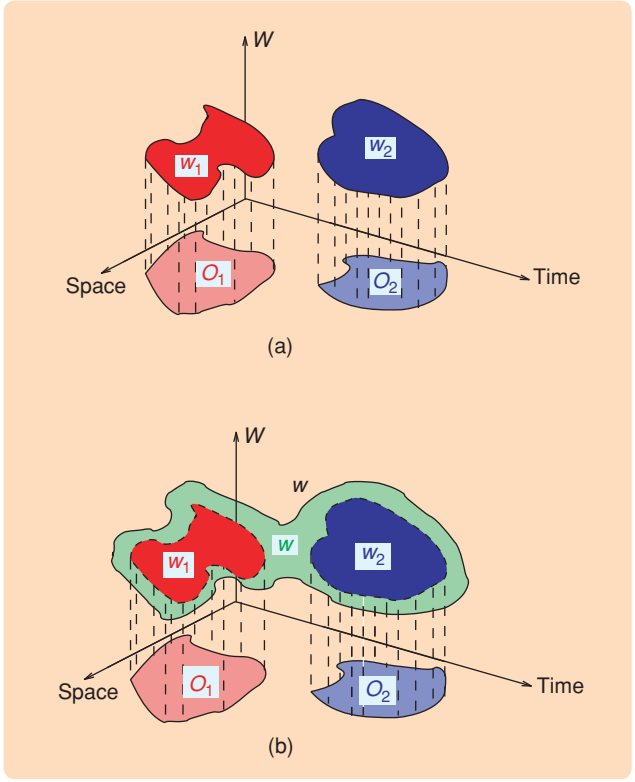


FIGURE 14 Controllability of systems described by PDEs. Behavioral controllability illustrated in Figure 13 generalizes seamlessly from 1D systems to nD systems. Parts (a) and (b) illustrate controllability for systems with many independent variables, for example, spatially distributed systems described by PDEs, where the independent variables include both time and space. For such systems controllability requires that any two patches of elements of the behavior, shown in (a), can be seen as patches of one and the same element in the behavior. This patchability is illustrated in (b).

The Initial Value Problem

For input/state/output models

$$\frac{d}{dt}x(t) = f(x(t), u(t)), \quad y(t) = h(x(t), u(t)), \tag{S1}$$

it is evident how elements of the behavior are generated. Roughly speaking, for any input $u : \mathbb{R} \to \mathbb{U}$ and initial condition $x(0) = x_0 \in \mathbb{X}$, there exists a unique solution $x : \mathbb{R} \to \mathbb{X}$ to (S1) and a unique output $y : \mathbb{R} \to \mathbb{Y}$. Rigorously, however, the vector field f and the input u must satisfy suitable smoothness conditions for the existence of a unique solution. The external behavior of this system, that is, the family of input/output pairs (u, y) defined by (S1), is basically parameterized by a free initial state $x(0)$ and a free input u.

Input/state/output models explicitly show how to parameterize the behavior. Other representations, such as image representations [see (20) and (29)], parameterize the behavior by means of a free exogenous latent input without intervention of initial conditions. However, for image representations the free exogenous inputs are not components of the vector of manifest variables.

In the context of linear time-invariant differential systems (LTIDSs), system (S1) becomes

$$\frac{d}{dt}x = Ax + Bu, \quad y = Cx + Du, \tag{S2}$$

where $u : \mathbb{R} \to \mathbb{R}^{\mathtt{m}}$, $x : \mathbb{R} \to \mathbb{R}^{\mathtt{n}}$, $y : \mathbb{R} \to \mathbb{R}^{\mathtt{p}}$, and A, B, C, D are real matrices of suitable dimension. For all $u \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{m}})$ and $x_0 \in \mathbb{R}^{\mathtt{n}}$, (S2) has a unique solution $x \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{n}})$ satisfying $x(0) = x_0$ and a unique corresponding $y \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{p}})$. Of course, one can get by with less smoothness than $\mathcal{C}^\infty$, for example, $u \in \mathcal{L}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{m}})$, and obtain a unique solution $x \in \mathcal{L}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{n}})$ (in fact, x is absolutely continuous) and $y \in \mathcal{L}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{p}})$, but, for simplicity, we consider only $\mathcal{C}^\infty$ trajectories, since this assumption is of no consequence for the algorithms given below on the construction of state representations.

DRIVING-VARIABLE AND OUTPUT-NULLING REPRESENTATIONS

The state representations (S1) and (S2) show how the system trajectories of the dynamical system are generated by a free input and a free initial state. However, alternative state representations are available. For example, the *driving-variable representation*

$$\begin{aligned} \frac{dx}{dt} &= Ax + Bv, \\ w &= Cx + Dv, \end{aligned}$$

in which the system trajectory $w : \mathbb{R} \to \mathbb{R}^{\mathtt{w}}$ is generated by a free driving variable $v : \mathbb{R} \to \mathbb{R}^{\mathtt{v}}$ and a free initial state. A driving-variable representation, which is the state analogue of an image representation, gives an explicit parameterization of the behavior. However, contrary to image representations, which represent only controllable LTIDSs, driving-variable representations can represent any LTIDS, including those that are not controllable.

Another state representation of LTIDSs is the *output-nulling* representation

$$\begin{aligned} \frac{dx}{dt} &= Ax + Bw, \\ 0 &= Cx + Dw, \end{aligned}$$

which can be viewed as the state analogue of a kernal representation. Driving-variable representations, as well as output-nulling representations, have input/state/output representations as a special case and are special cases of the first-order *differential-algebraic system* (DAE)

$$E\frac{dx}{dt} = Ax + Bw.$$

First-order DAEs express the external w-behavior by means of a differential equation that is first order in the latent variable x and zeroth order in the manifest variable w. It can be shown [2], [25] that the fact that the differential equation governing the behavior of (w, x) is first order in the latent variable and zeroth order in the manifest variable characterizes the state property of a latent variable.

THE STATE MAP

The set of input/output pairs $(u, y) \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{m}+\mathtt{p}})$ obtained this way is the *external behavior* of (S2), and (S3) is a *state representation* of its external behavior. The system (S2) is *minimal* if every system $(d/dt)x' = A'x' + B'u$, $y = C'x' + D'u$ with the same external behavior as (S2) has a state-space dimension greater than or equal to the state-space dimension of (S2). It is easy to prove that (S2) is minimal if and only if x is observable from (u, y), that is, if and only if

$$\mathtt{rank}\left(\begin{bmatrix} \mathcal{C}^\top A^\top C^\top \cdots (A^\top)^{\mathtt{dimension}(x)-1} C^\top \end{bmatrix}\right) = \mathtt{dimension}(x).$$

Note that, contrary to the classical state-space theory, controllability is not a requirement for minimality. The observable system $(d/dt)x = Ax$, $y = Cx$ is minimal in our sense. For example, the observable single-output system $(d/dt)x = Ax$, $y = Cx$ is a minimal state representation of the set of solutions of $p(d/dt)y = 0$, where $p(\xi) = \det(I\xi - A)$.

If follows from the elimination theorem (see the section "The Elimination Theorem for LTIDSs") that the external behavior of (S2) is an LTIDS, that is, there exists a polynomial matrix $R \in \mathbb{R}[\xi]^{\bullet \times (\mathtt{m}+\mathtt{p})}$ such that the external behavior of (S2) consists of the $\mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{m}+\mathtt{p}})$-solutions of $R(d/dt)\left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] = 0$. Minimality of (S2) implies, since it is equivalent to observability, that there exists a polynomial matrix $X \in \mathbb{R}[\xi]^{\mathtt{n} \times (\mathtt{m}+\mathtt{p})}$ such that, if (u, y, x) satisfies (S2), then $x = X(d/dt)\left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right]$. Hence, if (S2) is minimal and (u, y, x) is a solution of (S2), then

$$R\left(\frac{d}{dt}\right)\begin{bmatrix} u \\ y \end{bmatrix} = 0, \quad x = X\left(\frac{d}{dt}\right)\begin{bmatrix} u \\ y \end{bmatrix}.$$

The differential operator $X(d/dt)$ is a *minimal state map* [24] for $R(d/dt)\left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] = 0$. It is easy to see that the minimal state maps X_1, X_2 of two linear systems (S2) with the same external behavior are related by $X_1 \mapsto X_2 = SX_1$, with S a nonsingular matrix.

STATE CONSTRUCTION FROM A KERNEL REPRESENTATION

For systems that are not in input/state/output form, it is much less evident how to generate elements of the behavior. We now discuss this parameterization problem for LTIDSs. Consider the kernel representation (see the section "Linear Time-Invariant Systems")

$$R\left(\frac{d}{dt}\right)w = 0, \tag{S3}$$

where $R \in \mathbb{R}[\xi]^{\bullet\times \mathtt{W}}$. The behavior of (S3) is $\mathcal{B} = \text{kernel}\,(R(d/dt))$.

A matrix with one entry 1 in each row, at most one entry 1 in each column, and all other entries 0 is a *selector* matrix. The selector matrices S and S' are *complementary* if $\left[\begin{smallmatrix} S \\ S' \end{smallmatrix}\right]$ is a permutation matrix. Acting on a vector, a selector matrix selects out certain components of that vector, while a complementary selector matrix selects out the remaining components.

What are the free inputs and the free initial conditions that generate the elements of the behavior $\mathcal{B}$ of (S3)? We answer this question by constructing, from R, a selector matrix $S \in \mathbb{R}^{\bullet\times \mathtt{W}}$ and a polynomial matrix $X \in \mathbb{R}[\xi]^{\bullet\times \mathtt{W}}$, such that, for all $f \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet)$ and for all $x_0 \in \mathbb{R}^\bullet$, there exists a unique solution $w \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{W}})$ of (S3) such that $Sw = f$ and $X(d/dt)w(0) = x_0$. In other words, Sw and $X(d/dt)w(0)$ are free and specify $w \in \mathcal{B}$ uniquely. Equivalently, we construct a minimal input/state/output system (S2) with external behavior $\mathtt{kernel}(R(d/dt))$, the correspondence between solutions of (S2) and (S3) being $u = Sw$, $y = S'w$, and $x = X(d/dt)w$, with S' a selector matrix complementary to S. The algorithm given below is discussed in [1], [24], and [25], where more details can be found. See also [27].

Shifts of Polynomial Matrices

The central idea in the state construction algorithm is the *shift-and-cut map* $\sigma_- : \mathbb{R}[\xi]^{\bullet\times\bullet} \to \mathbb{R}[\xi]^{\bullet\times\bullet}$ that acts on polynomial matrices in the same way the backward shift acts on time functions. The action of σ_- on $P(\xi) = P_0 + P_1\xi + P_2\xi^2 + \cdots + P_{\mathtt{L}}\xi^{\mathtt{L}}$ is given by

$$\begin{aligned}\sigma_- : P_0 + P_1\xi + P_2\xi^2 + \cdots + P_{\mathtt{L}}\xi^{\mathtt{L}} \\ \mapsto P_1 + P_2\xi + \cdots + P_{\mathtt{L}}\xi^{\mathtt{L}-1}.\end{aligned}$$

Repeatedly acting with σ_- and stacking the resulting polynomial matrices leads to

$$\begin{aligned}\Sigma_-(P) &:= \begin{bmatrix} \sigma_-(P) \\ \sigma_-^2(P) \\ \vdots \\ \sigma_-^{\mathtt{L}-1}(P) \\ \sigma_-^{\mathtt{L}}(P) \end{bmatrix} \\ &= \begin{bmatrix} P_1 + P_2\xi + \cdots + P_{\mathtt{L}-1}\xi^{\mathtt{L}-2} + P_{\mathtt{L}}\xi^{\mathtt{L}-1} \\ P_2 + P_3\xi + \cdots + P_{\mathtt{L}}\xi^{\mathtt{L}-2} \\ \vdots \\ P_{\mathtt{L}-1} + P_{\mathtt{L}}\xi \\ P_{\mathtt{L}} \end{bmatrix}.\end{aligned}$$

Multiplication by ξ, on the other hand, which is denoted by σ_+, acts in the same way the forward shift acts on time functions. Its action on P is given by

$$\begin{aligned}\sigma_+ : P_0 + P_1\xi + P_2\xi^2 + \cdots + P_{\mathtt{L}}\xi^{\mathtt{L}} \\ \mapsto P_0\xi + P_1\xi^2 + P_2\xi^3 + \cdots + P_{\mathtt{L}}\xi^{\mathtt{L}+1}.\end{aligned}$$

Note that $\sigma_-\sigma_+ =$ identity, while $\sigma_+(\sigma_-(P)) = P - P(0)$. Consequently,

$$\sigma_+\left(\begin{bmatrix} \Sigma_-(P) \\ 0 \end{bmatrix}\right) = \begin{bmatrix} P \\ \Sigma_-(P) \end{bmatrix} - \begin{bmatrix} P_0 \\ P_1 \\ \vdots \\ P_{\mathtt{L}} \end{bmatrix}. \tag{S4}$$

Denote by $\mathtt{rowspan}_{\mathbb{R}}(P)$ the subspace spanned by the rows of $P \in \mathbb{R}[\xi]^{\bullet\times\bullet}$, viewed as elements of an $\mathbb{R}$-vector space of polynomial row vectors, and by $\mathtt{module}_{\mathbb{R}[\xi]}(P)$ the $\mathbb{R}[\xi]$-module spanned by the rows of P, that is, $g \in \mathtt{module}_{\mathbb{R}[\xi]}(P)$ if and only if there exists $f \in \mathbb{R}[\xi]^{1\times\bullet}$ such that $g = fP$. Equivalently,

$$\mathtt{module}_{\mathbb{R}[\xi]}(P) = \mathtt{rowspan}_{\mathbb{R}}\left(\begin{bmatrix} P \\ \sigma_+(P) \\ \sigma_+^2(P) \\ \vdots \end{bmatrix}\right).$$

Construction of a Minimal State Map

The operator Σ_- applied to the polynomial matrix R is the key to the construction of a minimal state map for (S3). Let $X \in \mathbb{R}[\xi]^{\bullet\times\mathtt{W}}$ be a polynomial matrix with independent rows, such that $\mathtt{rowspan}_{\mathbb{R}}(X)$ is a direct summand of $\mathtt{module}_{\mathbb{R}[\xi]}(R)$ relative to $\mathtt{rowspan}_{\mathbb{R}}(\Sigma_-(R)) + \mathtt{module}_{\mathbb{R}[\xi]}(R)$. In other words, X has independent rows and is such that

$$\begin{aligned}\mathtt{rowspan}_{\mathbb{R}}(X) \oplus \mathtt{module}_{\mathbb{R}[\xi]}(R) = \\ \mathtt{rowspan}_{\mathbb{R}}(\Sigma_+(R)) + \mathtt{module}_{\mathbb{R}[\xi]}(R).\end{aligned}$$

The differential operator $X(d/dt)$ is a minimal state map for (S3).

The computation of the matrix $X(\xi) = X_0 + X_1\xi + \cdots + X_{\mathtt{degree}(X)}\xi^{\mathtt{degree}(X)}$ from $R(\xi) = R_0 + R_1\xi + \cdots + R_{\mathtt{degree}(R)}\xi^{\mathtt{degree}(R)}$ can be carried out by analyzing a Hankel matrix. The construction of X requires that the row span of the first matrix shown at the bottom of the next page is equal to the row span of the second matrix at the bottom of the next page and that the rows of the last block row

$$[\,X_0 \quad X_1 \quad X_2 \quad \cdots \quad X_{\mathtt{degree}(X)} \quad 0 \quad \cdots\,]$$

are linearly independent as well as linearly independent of the rows above it.

Construction of the Input Selector Matrix

The input and output components of (S3) are obtained by choosing a selector matrix $S \in \mathbb{R}^{\bullet\times\mathtt{W}}$ such that $\mathtt{rowspan}_{\mathbb{R}}(S)$ is a direct summand for $\mathtt{rowspan}_{\mathbb{R}}(\Sigma_-(R)) + \mathtt{module}_{\mathbb{R}[\xi]}(R)$ relative to $\mathtt{rowspan}_{\mathbb{R}}(I_{\mathtt{W}\times\mathtt{W}}) + \mathtt{rowspan}_{\mathbb{R}}(\Sigma_-(R)) + \mathtt{module}_{\mathbb{R}[\xi]}(R)$, that is,

$$\begin{aligned}&\mathtt{rowspan}_{\mathbb{R}}(S) \oplus \big(\mathtt{rowspan}_{\mathbb{R}}(\Sigma_-(R)) + \mathtt{module}_{\mathbb{R}[\xi]}(R)\big)\\ &= \mathtt{rowspan}_{\mathbb{R}}(I_{\mathtt{w}\times\mathtt{w}}) + \mathtt{rowspan}_{\mathbb{R}}(\Sigma_-(R)) + \mathtt{module}_{\mathbb{R}[\xi]}(R).\end{aligned}$$

The construction of S can be carried out by first computing the subspace $\mathcal{L}_0 \subseteq \mathbb{R}^{1\times \mathtt{w}}$ consisting of the $\mathbb{R}$-span of the elements with degree zero in $\mathtt{rowspan}_{\mathbb{R}}(\Sigma_-(R)) + \mathtt{module}_{\mathbb{R}[\xi]}(R)$ and choosing S such that the rows of S span a complement of $\mathcal{L}_0$ relative to $\mathbb{R}^{\mathtt{w}}$, viewed as the space of $\mathtt{w}$-dimensional row vectors. Let S' be a selector matrix that is complementary to S. The trajectory Sw functions as input, while $S'w$ functions as output for (S3).

Computation of the System Matrices

The matrices X and S jointly lead to the direct sum decomposition

$$\begin{aligned}&\mathtt{rowspan}_{\mathbb{R}}(I_{\mathtt{w}\times\mathtt{w}}) + \mathtt{rowspan}_{\mathbb{R}}(\Sigma_-(R)) + \mathtt{module}_{\mathbb{R}[\xi]}(R)\\ &\quad = \mathtt{rowspan}_{\mathbb{R}}(S) \oplus \mathtt{rowspan}_{\mathbb{R}}(X) \oplus \mathtt{module}_{\mathbb{R}[\xi]}(R).\end{aligned}$$

From (S4) it follows that

$$\begin{aligned}\mathtt{rowspan}_{\mathbb{R}}(\sigma_+(\Sigma_-(R))) &\subseteq \mathtt{rowspan}_{\mathbb{R}}(\Sigma_-(R))\\ &\quad + \mathtt{module}_{\mathbb{R}[\xi]}(R) + \mathtt{rowspan}_{\mathbb{R}}(I_{\mathtt{w}\times\mathtt{w}}),\end{aligned}$$

which implies

$$\begin{aligned}\mathtt{rowspan}_{\mathbb{R}}(\sigma_+(X)) &\subseteq \mathtt{rowspan}_{\mathbb{R}}(X)\\ &\quad \oplus \mathtt{rowspan}_{\mathbb{R}}(S) \oplus \mathtt{module}_{\mathbb{R}[\xi]}(R).\end{aligned}$$

The construction of S' yields

$$\begin{aligned}\mathtt{rowspan}_{\mathbb{R}}(S') &\subseteq \mathtt{rowspan}_{\mathbb{R}}(X)\\ &\quad \oplus \mathtt{rowspan}_{\mathbb{R}}(S) \oplus \mathtt{module}_{\mathbb{R}[\xi]}(R).\end{aligned}$$

These inclusions imply that there exist unique matrices A, B, C, D and polynomial matrices E, F such that

$$\xi X(\xi) = AX(\xi) + BS + E(\xi)R(\xi), \tag{S5}$$

$$S' = CX(\xi) + DS + F(\xi)R(\xi). \tag{S6}$$

The system

$$\frac{d}{dt}x = Ax + Bu, \quad y = Cx + Du, \quad w = \begin{bmatrix} S \\ S' \end{bmatrix}^{-1} \begin{bmatrix} u \\ y \end{bmatrix}$$

is a minimal state representation of (S3). The trajectory w in the behavior $\mathcal{B}$ of (S3) corresponds to the state trajectory $x = X(d/dt)w$, input trajectory $u = Sw$, and output trajectory $y = S'w$. This input/state/output representation puts in evidence how elements of $\mathcal{B}$ are generated. For every vector $\mathrm{x}_0 \in \mathbb{R}^{\mathtt{n}}$ with $\mathtt{n}$ equal to the number of rows of X, and for every $u \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{m}})$ with $\mathtt{m}$ equal to the number of rows of S, there exists a unique element $w \in \mathcal{B}$ that satisfies $X(d/dt)w(0) = \mathrm{x}_0$ and $Sw = u$.

Note that controllability plays no role in this algorithm for constructing a minimal input/state/output representation of a system in kernel representation. Since a minimal state representation (S2) is state controllable if and only if the external behavior is controllable in the behavioral sense, the minimal state construction yields a state controllable representation if and only if the system (S3) is controllable in the behavioral sense, that is, if and only if $\mathtt{module}_{\mathbb{R}[\xi]}(R)$ is a closed module (see "Polynomial Modules and Syzygies").

In [25] this algorithm for constructing an input/state/output representation of the behavior of an LTIDS system in kernel representation is adapted to systems with latent variables or in image representation.

SINGLE-INPUT/SINGLE-OUTPUT SYSTEMS

The algorithm of the previous section gives insight into the derivation of classical state-space representations for single-input/single-output system described by

$$\begin{bmatrix}
 & & \vdots & & & & & & \\
0 & 0 & R_0 & & & \cdots & & R_{\mathtt{degree}(R)} & 0 \ \cdots \\
0 & R_0 & R_1 & & \cdots & & R_{\mathtt{degree}(R)} & 0 & \cdots \\
R_0 & R_1 & R_2 & \cdots & & R_{\mathtt{degree}(R)} & 0 & \cdots & \\
R_1 & R_2 & & & R_{\mathtt{degree}(R)} & 0 & \cdots & & \\
R_2 & & \cdots & R_{\mathtt{degree}(R)} & 0 & \cdots & & & \\
 & & \vdots & & & & & & \\
R_{\mathtt{degree}(R)-1} & R_{\mathtt{degree}(R)} & 0 \ \cdots & & & & & & \\
R_{\mathtt{degree}(R)} & 0 & \cdots & & & & & &
\end{bmatrix}$$

$$\begin{bmatrix}
 & & \vdots & & & & \\
0 & 0 & R_0 & & \cdots & R_{\mathtt{degree}(R)} & 0 \ \cdots \\
0 & R_0 & R_1 & \cdots & R_{\mathtt{degree}(R)} & 0 & \cdots \\
R_0 & R_1 & R_2 \ \cdots & R_{\mathtt{degree}(R)} & 0 & \cdots & \\
X_0 & X_1 & X_2 \ \cdots & X_{\mathtt{degree}(X)} & 0 & \cdots &
\end{bmatrix},$$

$$p\left(\frac{d}{dt}\right)y = q\left(\frac{d}{dt}\right)u, \tag{S7}$$

where $p, q \in \mathbb{R}[\xi]$,

$$p(\xi) = p_0 + p_1\xi + \cdots + p_{\mathrm{n}-1}\xi^{\mathrm{n}-1} + p_{\mathrm{n}}\xi^{\mathrm{n}},$$
$$q(\xi) = q_0 + q_1\xi + \cdots + q_{\mathrm{n}-1}\xi^{\mathrm{n}-1} + q_{\mathrm{n}}\xi^{\mathrm{n}}.$$

Assume that the degree of p is greater than or equal to the degree of p. For simplicity, we take $p_{\mathrm{n}} = 1$.

The aim of this section is to show that the application of the shift-and-cut operator leads in a straightforward way to the standard state-space representations. The shift-and-cut operator applied to (S7) yields the polynomial matrix

$$\Sigma_-([p \quad -q])(\xi) = \begin{bmatrix} p_1 + p_2\xi + \cdots + p_{\mathrm{n}}\xi^{\mathrm{n}-1} & -q_1 - q_2\xi - \cdots - q_{\mathrm{n}}\xi^{\mathrm{n}-1} \\ p_2 + \cdots + p_{\mathrm{n}}\xi^{\mathrm{n}-2} & -q_2 - \cdots - p_{\mathrm{n}}\xi^{\mathrm{n}-2} \\ \vdots & \vdots \\ p_{\mathrm{n}-1} + p_{\mathrm{n}}\xi & -q_{\mathrm{n}-1} - q_{\mathrm{n}}\xi \\ p_{\mathrm{n}} & -q_{\mathrm{n}} \end{bmatrix}. \tag{S8}$$

This $\mathrm{n} \times 2$ matrix has $\mathbb{R}$-linearly independent rows and their span is $\mathbb{R}$-linearly independent of $\mathtt{module}_{\mathbb{R}[\xi]}([p \quad -q])$. The construction of a minimal state map requires choosing a polynomial matrix $X \in \mathbb{R}[\xi]^{\mathrm{n}\times 2}$ such that $\mathtt{rowspan}_{\mathbb{R}}(X)$ equals the rowspan of (S8). For the case at hand, since (S8) has $\mathbb{R}$-independent rows, no selection of rows of $\Sigma_-([p \quad -q])(\xi)$ is needed to construct the polynomial matrix X.

The construction of the selector matrix S is also evident for the case at hand, since it requires that the rows of $\left[\begin{smallmatrix} p_{\mathrm{n}} & -q_{\mathrm{n}} \\ S & \end{smallmatrix}\right]$ span the rows of $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, leading to $S = [0 \quad 1]$ and $S' = [1 \quad 0]$.

There are two convenient choices for X that lead to input/state/output representations. The first choice consists of the rows of (S8) in reverse order,

$$X(\xi) = \begin{bmatrix} p_{\mathrm{n}} & -q_{\mathrm{n}} \\ p_{\mathrm{n}-1} + p_{\mathrm{n}}\xi & -q_{\mathrm{n}-1} - q_{\mathrm{n}}\xi \\ \vdots & \vdots \\ p_2 + \cdots + p_{\mathrm{n}}\xi^{\mathrm{n}-2} & -q_2 - \cdots - p_{\mathrm{n}}\xi^{\mathrm{n}-2} \\ p_1 + p_2\xi + \cdots + p_{\mathrm{n}}\xi^{\mathrm{n}-1} & -q_1 - q_2\xi - \cdots - q_{\mathrm{n}}\xi^{\mathrm{n}-1} \end{bmatrix}.$$

Applying algorithm (S5), (S6) with this choice of X yields

$$A = \begin{bmatrix} -p_{\mathrm{n}-1} & 1 & 0 & \cdots & 0 \\ -p_{\mathrm{n}-2} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -p_1 & 0 & 0 & \cdots & 1 \\ -p_0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad B = \begin{bmatrix} q_{\mathrm{n}-1} - p_{\mathrm{n}-1}q_{\mathrm{n}} \\ q_{\mathrm{n}-2} - p_{\mathrm{n}-2}q_{\mathrm{n}} \\ \vdots \\ q_1 - p_1 q_{\mathrm{n}} \\ q_0 - p_0 q_{\mathrm{n}} \end{bmatrix},$$

$$C = [1 \quad 0 \quad 0 \quad \cdots \quad 0], \qquad D = q_{\mathrm{n}}.$$

This input/state/output representation of (S7) is the *observer canonical form* [S7, p. 42].

The second convenient choice for X is

$$X(\xi) = \begin{bmatrix} 1 & b_0 \\ \xi & b_1 + b_0\xi \\ \vdots & \vdots \\ \xi^{\mathrm{n}-2} & b_{\mathrm{n}-2} + \cdots + b_0\xi^{\mathrm{n}-2} \\ \xi^{\mathrm{n}-1} & b_{\mathrm{n}-1} + b_{\mathrm{n}-2}\xi + \cdots + b_0\xi^{\mathrm{n}-1} \end{bmatrix}.$$

The coefficients $b_0, b_1, \ldots, b_{\mathrm{n}-1}, b_{\mathrm{n}}$ are enforced by the special form of the first column of X and the requirement that the rows of X span the rows of (S8). These coefficients can be computed from the expansion

$$\frac{q(\xi)}{p(\xi)} = b_0 + b_1\xi^{-1} + b_2\xi^{-2} + \cdots + b_{\mathrm{n}-1}\xi^{-\mathrm{n}+1} + b_{\mathrm{n}}\xi^{-\mathrm{n}} + \cdots. \tag{S9}$$

Applying algorithm (S5), (S6) with this choice of X yields

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -p_0 & -p_1 & -p_2 & \cdots & -p_{\mathrm{n}-1} \end{bmatrix}, \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{\mathrm{n}-1} \\ b_{\mathrm{n}} \end{bmatrix},$$

$$C = [1 \quad 0 \quad 0 \quad \cdots \quad 0], \qquad D = b_0.$$

This input/state/output representation of (S7) is the *observable canonical form* [S7, p. 43].

Both the observer and the observable canonical form are observable, hence minimal, state representations of (S7), and state controllable if and only if (S7) is controllable in the behavioral sense, that is, if and only if p and q are coprime.

The state construction algorithm can also be applied to obtain a *modal canonical form*. Assume, for simplicity, that we work over $\mathbb{C}$ and that $p_0 + p_1\xi + \cdots + p_{\mathrm{n}-1}\xi^{\mathrm{n}-1} + p_{\mathrm{n}}\xi^{\mathrm{n}} \in \mathbb{C}[\xi]$ has distinct roots $\lambda_1, \lambda_2, \ldots, \lambda_{\mathrm{n}} \in \mathbb{C}$. Choose

$$X(\xi) = \begin{bmatrix} \frac{p(\xi)}{\xi - \lambda_1} & -\frac{q(\xi)}{\xi - \lambda_1} \\ \frac{p(\xi)}{\xi - \lambda_2} & -\frac{q(\xi)}{\xi - \lambda_2} \\ \vdots & \vdots \\ \frac{p(\xi)}{\xi - \lambda_{\mathrm{n}}} & -\frac{q(\xi)}{\xi - \lambda_{\mathrm{n}}} \end{bmatrix}.$$

The rows of this matrix X span the rows of (S8), since for $\lambda \in \mathbb{C}$ and $r \in \mathbb{C}[\xi]$, $r(\xi) = r_0 + r_1\xi + \cdots + r_{\mathrm{L}}\xi^{\mathrm{L}}$, the following relation holds:

$$\frac{r(\xi) - r(\lambda)}{\xi - \lambda} = [1 \quad \lambda \quad \cdots \quad \lambda^{\mathrm{L}-2} \quad \lambda^{\mathrm{L}-1}] \times \begin{bmatrix} r_1 + \cdots + r_{\mathrm{L}-1}\xi^{\mathrm{L}-2} + r_{\mathrm{L}}\xi^{\mathrm{L}-1} \\ r_2 + \cdots + r_{\mathrm{L}}\xi^{\mathrm{L}-2} \\ \vdots \\ r_{\mathrm{L}-1} + r_{\mathrm{L}}\xi \\ r_{\mathrm{L}} \end{bmatrix}.$$

Hence X is equal to (S8) premultiplied by a Vandermonde matrix.

Applying algorithm (S5), (S6) with this choice of X yields the equations

$$\xi X(\xi) = AX(\xi) + B[\,0 \quad 1\,] + \begin{bmatrix} \gamma_1(\xi) \\ \gamma_2(\xi) \\ \vdots \\ \gamma_{\mathtt{n}}(\xi) \end{bmatrix} [\,p(\xi) \quad -q(\xi)\,],$$

$$[\,0 \quad 1\,] = CX(\xi) + D[\,0 \quad 1\,] + \gamma_0(\xi)\ [\,p(\xi) \quad -q(\xi)\,].$$

These equations yield $\gamma_0 = 0$ and $\gamma_1 = \gamma_2 = \cdots = \gamma_{\mathtt{n}} = 1$, while A, B, C, D are given by

$$A = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{\mathtt{n}} \end{bmatrix}, \qquad B = \begin{bmatrix} q(\lambda_1) \\ q(\lambda_2) \\ \vdots \\ q(\lambda_{\mathtt{n}}) \end{bmatrix},$$

$$C = [\,\tfrac{1}{\pi(\lambda_1)} \quad \tfrac{1}{\pi(\lambda_2)} \quad \cdots \quad \tfrac{1}{\pi(\lambda_{\mathtt{n}})}\,], \qquad D = q_{\mathtt{n}}.$$

This input/state/output system is state controllable if and only if $q(\lambda_{\mathtt{k}}) \neq 0$ for $\mathtt{k} = 1, 2, \ldots, \mathtt{n}$. Equivalently, if and only if p and q are coprime, which is the condition for the behavioral controllability of (S7).

State Construction from an Image Representation of (S7)

When p and q are coprime, (S7) admits the image representation

$$\begin{bmatrix} u \\ y \end{bmatrix} = M\left(\frac{d}{dt}\right)\ell, \quad M = \begin{bmatrix} p \\ q \end{bmatrix}. \tag{S10}$$

Applying the algorithm for state construction on (S10) with system variables (ℓ, u, y) leads to

$$\Sigma_-([\,M \quad I_{2\times 2}\,]) = [\,\Sigma_-(M) \quad 0 \quad 0\,],$$

with

$$\Sigma_-(M)(\xi) = \begin{bmatrix} p_1 + p_2\xi + \cdots + p_{\mathtt{n}}\xi^{\mathtt{n}-1} \\ q_1 + q_2\xi - \cdots - q_{\mathtt{n}}\xi^{\mathtt{n}-1} \\ p_2 + \cdots + p_{\mathtt{n}}\xi^{\mathtt{n}-2} \\ q_2 + \cdots + p_{\mathtt{n}}\xi^{\mathtt{n}-2} \\ \vdots \\ p_{\mathtt{n}-1} + p_{\mathtt{n}}\xi \\ q_{\mathtt{n}-1} + q_{\mathtt{n}}\xi \\ p_{\mathtt{n}} \\ q_{\mathtt{n}} \end{bmatrix}. \tag{S11}$$

A minimal state for (S10) is therefore given by $x = X(d/dt)\ell$ with X a polynomial vector with $\mathbb{R}$-independent rows such that $\mathtt{rowspan}_{\mathbb{R}}(X)$ equals the $\mathbb{R}$-span of the rows of (S11). The rank of (S11) is $\mathtt{n}$, and hence $X \in \mathbb{R}[\xi]^{\mathtt{n}}$.The input is given by $u = SM(d/dt)\ell$ with S a suitable selector matrix. The input selector matrix S is chosen such that $\mathtt{rowspan}_R(X)$ and $\mathtt{rowspan}_{\mathbb{R}}(SM)$ are direct summands. Let S' be a complementary selector matrix. As expected, $S = [\,1 \quad 0\,]$ and $S' = [\,0 \quad 1\,]$ satisfy these requirements.

It is easily seen that

$$\mathtt{rowspan}_{\mathbb{R}}(\sigma_+(X)) \subseteq \mathtt{rowspan}_R(X) \oplus \mathtt{rowspan}_{\mathbb{R}}(SM)$$

and

$$\mathtt{rowspan}_{\mathbb{R}}(S'M) \subseteq \mathtt{rowspan}_R(X) \oplus \mathtt{rowspan}_{\mathbb{R}}(SM).$$

From these inclusions, it follows that the equations

$$\xi X(\xi) = AX(\xi) + BSM(\xi), \tag{S12}$$
$$S'M(\xi) = CX(\xi) + DSM(\xi) \tag{S13}$$

have unique solutions A, B, C, D. The matrices A, B, C, D define an input/state/output representation of (S10). The corresponding input, output, and state trajectories are given by $u = SM(d/dt)\ell$, $y = S'M(d/dt)\ell$, and $x = X(d/dt)\ell$.

There are again two convenient choices for $X \in \mathbb{R}[\xi]^{\mathtt{n}}$. The first choice is

$$X(\xi) = \begin{bmatrix} p_{\mathtt{n}} \\ p_{\mathtt{n}-1} + p_{\mathtt{n}}\xi \\ \vdots \\ p_2 + \cdots + p_{\mathtt{n}}\xi^{\mathtt{n}-2} \\ p_1 + p_2\xi + \cdots + p_{\mathtt{n}}\xi^{\mathtt{n}-1} \end{bmatrix}.$$

Applying algorithm (S12), (S13) yields

$$A = \begin{bmatrix} -p_{\mathtt{n}-1} & 1 & 0 & \cdots & 0 \\ -p_{\mathtt{n}-2} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -p_1 & 0 & 0 & \cdots & 1 \\ -p_0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

$$C = [\,b_1 \quad b_2 \quad b_3 \cdots b_{\mathtt{n}}\,], \qquad D = b_0.$$

The coefficients $b_0, b_1, \ldots, b_{\mathtt{n}}$ are computed by the expansion (S9). This input/state/output representation of (S7) is the *controllable canonical form* [S7, p. 43].

The second choice for X is

$$X(\xi) = \begin{bmatrix} 1 \\ \xi \\ \vdots \\ \xi^{\mathtt{n}-2} \\ \xi^{\mathtt{n}-1} \end{bmatrix}.$$

Applying algorithm (S12), (S13) yields

$$(w_1, w_2'), \quad (w_1, w_2'') \in \mathcal{B}$$

implies

$$w_1' = w_2'',$$

that is, if there exists a map $F : \mathbb{W}_1^{\mathbb{T}} \to \mathbb{W}_2^{\mathbb{T}}$ such that

$$(w_1, w_2) \in \mathcal{B}$$

implies

$$w_2 = F(w_1).$$

Observability means that, by observing the trajectory $w_1 : \mathbb{T} \to \mathbb{W}_1$ and knowing the laws that govern the

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -p_0 & -p_1 & -p_2 & \cdots & -p_{\mathtt{n}-1} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

$$C = [\, q_0 - p_0 q_{\mathtt{n}} \quad q_1 - p_1 q_{\mathtt{n}} \quad \cdots \quad q_{\mathtt{n}-1} - p_{\mathtt{n}-1} q_{\mathtt{n}} \,], \; D = q_{\mathtt{n}}.$$

This input/state/output representation of (S7) is the *controller canonical form* [S7, p. 39].

Both the controllable and the controller canonical form are state controllable and state observable, hence minimal, input/state/output representations of (S7), assuming that p and q are coprime. The algorithm (S12), (S13) can be carried out even when p and q are not coprime. In this case, the controllable and the controller canonical form are state controllable but not state observable, hence nonminimal, input/state/output representations of the behavior of the image representation (S10) and of the controllable part of the kernel representation (S7).

SYSTEMS WITH TIME-AXIS $[0, \infty)$

Considering (S3) as a system defined on the time-axis $[0, \infty)$ presents no difficulties, provided, of course, that derivatives at $t = 0$ are taken to be one-sided derivatives from the right. Denote the $\mathtt{n}$-th one-sided derivative of w from the right at 0 by $(d^{\mathtt{n}}/dt^{\mathtt{n}})w(0^+)$. Let $\mathcal{B}_+$ be the set of $\mathcal{C}^\infty([0, \infty), \mathbb{R}^{\mathtt{w}})$-solutions of (S3). The dynamical system $([0, \infty), \mathbb{R}^{\mathtt{w}}, \mathcal{B}_+)$ is linear and time invariant since $\sigma^t(\mathcal{B}_+) \subseteq \mathcal{B}_+$ for all $t \in [0, \infty)$. In fact, $\mathcal{B}_+ = \mathcal{B}|_{[0,\infty)}$, with $\mathcal{B}$ the set of $\mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{w}})$-solutions of (S3) and $\sigma^t(\mathcal{B}_+) = \mathcal{B}_+$ for all $t \in [0, \infty)$.

The state construction and input selection algorithms carry over verbatim and yield a selector matrix $S \in \mathbb{R}^{\bullet \times \mathtt{w}}$ and a polynomial matrix $X \in \mathbb{R}^{\bullet \times \mathtt{w}}$ such that, for all $f \in \mathcal{C}^\infty([0, \infty), \mathbb{R}^\bullet)$ and for all $\mathtt{x}_0 \in \mathbb{R}^\bullet$, there exists a unique $w \in \mathcal{B}_+$ such that $Sw = f$ and $X(d/dt)w(0^+) = \mathtt{x}_0$. A $\mathcal{C}^\infty([0, \infty), \mathbb{R}^{\mathtt{w}})$-solution of (S3) is uniquely specified by the input $Sw \in \mathcal{C}^\infty([0, \infty), \mathbb{R}^{\mathtt{w}})$ and the initial condition $X(d/dt)w(0^+)$.

For the example

$$y + \frac{d}{dt}y + \frac{d^2}{dt^2}y = u + \frac{d}{dt}u, \quad w = \begin{bmatrix} u \\ y \end{bmatrix},$$

the state map X can be taken to be

$$X(\xi) = \begin{bmatrix} 0 & 1 \\ -1 & 1 + \xi \end{bmatrix}.$$

It follows that, for all $u \in \mathcal{C}^\infty([0, \infty), \mathbb{R})$ and for all $\mathtt{x}_0 \in \mathbb{R}^2$, there exists a unique solution y to this differential equation on $[0, \infty)$ such that $X(d/dt)w(0^+) = \mathtt{x}_0$. The unique output that corresponds to $u(t) = e^{-t}$ for $t \in [0, \infty)$ and $(d/dt)y(0^+) = 0, y(0^+) = 0$, is $y(t) = 0$ for $t \in [0, \infty)$. The difficulties discussed in [S8] and [S9] that occur as a consequence of using Laplace transforms on $[0, \infty)$ are not encountered.

WELL-POSEDNESS

Often a model described in terms of differential equations is called *well-posed* [S10, p. 227] if the model has

1) existence of solutions
2) uniqueness of solutions
3) continuous dependence on parameters.

This principle is attributed to Hadamard. Existence and uniqueness refer to the fact that an ODE or a PDE is required to have a unique solution given initial conditions and boundary conditions. This requirement is strange since a differential equation model of a physical system usually does not come with an initial condition. As we have seen for (S3), the initial conditions that give existence and uniqueness of a solutions have to be constructed. For Kepler's laws, the differential equation whose solutions are the trajectories that satisfy Kepler's laws has to be constructed. The initial conditions that give existence and uniqueness for this second order differential equation consist of the initial position and velocity. To have existence and uniqueness for Maxwell's equations on a domain, we need to construct, from the equations, suitable initial and boundary conditions. The initial and boundary conditions are not physical laws. Nowadays, when referring to the continuous dependence on parameters that appears in the definition of well-posedness, one usually adds "in a suitable topology."

Hadamard's well-posedness principle thus requires existence, uniqueness, and continuous dependence on parameters if the initial conditions, the boundary conditions, and the topology are chosen so as to give existence, uniqueness, and continuous dependence on parameters. The principle is an example of circular reasoning.

REFERENCES

[S7] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice Hall, 1980.

[S8] K.H. Lundberg, H.R. Miller, and D.L. Trumper, "Initial conditions, generalized functions, and the Laplace transform," *IEEE Control Syst. Mag.*, vol. 27, no. 1, pp. 22–31, 2007.

[S9] K.H. Lundberg and D.S. Bernstein, "ODEs with differentiated inputs," *IEEE Control Syst. Mag.*, vol. 27, no. 3, pp. 95–97, 2007.

[S10] R. Courant and D. Hilbert, *Methods of Mathematical Physics*. New York: Wiley Interscience, 1953.

system $\mathcal{B}$, it is possible to deduce the unobserved trajectory $w_2 : \mathbb{T} \to \mathbb{W}_2$ uniquely. Finally, w_2 is *detectable* from w_1 if

$$(w_1, w_2'), (w_1, w_2'') \in \mathcal{B}$$

implies

$$w_2'(t) - w_2''(t) \to 0 \quad \text{as} \quad t \to \infty,$$

assuming that $\mathbb{W}_2$ is a subset of a normed vector space.

The definitions of controllability, stabilizability, observability, and detectability have the usual state-space definitions as special cases. These extensions to behaviors

A Brief History of Controllability

The introduction of controllability is one of the milestones in the history of control. This notion played a seminal role in the early development of the state-space theory of dynamical systems. Since then, controllability features as a regularizing assumption in essentially every major theoretical development in the field. By now, controllability is one of the first things taught in introductory control courses. A state-space system is *controllable* if, for any two states, there is an input that drives the system from the first state to the second. This definition is illustrated in Figure 13.

A good overview of the early history of controllability is given in the lecture notes [S11] by Rudy Kalman. After discussing the state-space notions of controllability, observability, and minimal state representation, it is stated that

> *Only much later did it become clear, however, that a dynamical system is always controllable if it is derived from an external description.* [S11, p. 136]

This remark refers to the result, obtained a few years after the introduction of the notions of controllability and observability, that every transfer function and impulse response can be represented by a minimal state representation, minimal in the sense that the state-space dimension is as small as possible and that minimality is equivalent to the combined properties of state controllability and state observability. The conviction that every system can be represented by a state-controllable and state-observable representation has had a deep influence in the field.

The validity of the quoted statement hinges, of course, on the meaning of the term "external description." If one considers this term to mean a transfer function or an impulse response, then the statement is perfectly correct. But transfer functions and impulse responses assume, in a somewhat disguised form, that the system starts in zero initial conditions. This assumption is rather limited, from a historical, theoretical, and practical point of view. Zeroing the initial conditions and identifying a system with its transfer function does not do justice to the work of Maxwell, Routh, Hurwitz, and Lyapunov, since this assumption excludes autonomous systems such as $p(d/dt)w = 0$ with p a real polynomial, as well as $(d/dt)x = Ax$, with A a square matrix. For many linear systems, such as Newton's second law, and for nonlinear systems, fixing the initial conditions is awkward. In nonlinear models, there is often no reason to prefer one initial condition above another one. In Kepler's laws, for example, fixing the initial conditions eliminates all trajectories but one.

That the classical concept of state controllability is merely a property of a representation, and not of the system itself, has been articulated forcefully by Charlie Desoer:

> *Some authors write "We assume that the transfer function $\hat{H}(s)$ is completely controllable and completely observable." This is sheer nonsense. […] The important point is that the concepts of observability and of controllability are properties of the representation and not of the model represented.* [S12, p. 181]

The question emerges as to whether there is a way to view controllability as a genuine system property, as a property of the external description, rather than of only a state-space representation. Kalman's statement suggests a negative answer, since viewed from an external description, every system appears to be controllable. It is one of the main contributions of the behavioral approach to have pinpointed controllability as an authentic system property. In this approach, a system is viewed as a family of trajectories, and controllability has to do with the way system trajectories are intertwined. A behavior is defined to be controllable if it is possible to transfer from any past trajectory to any future trajectory, while obeying the dynamical laws of the system (see Figure 13). This definition is applicable to nonlinear, discrete-event, and delay-differential systems without having to introduce a state representation, and can be generalized seamlessly to nD systems (see Figure 14), where the notion of state is generally unclear. Controllability for nD systems is discussed [45], [46] for discrete 2D systems and in [34] for PDEs. The development of the notion of controllability for nD systems is discussed in [S13].

For linear time-invariant differential systems, effective linear-algebra-based tests are available for behavioral controllability. Consider the system with behavior $\mathcal{B} = \ker(R(d/dt))$, where R is a real polynomial matrix. The goal is to derive tests on R for controllability of $\mathcal{B}$. The most useful tests are given by the equivalence of the following conditions:

1) $\mathcal{B} = \ker(R(d/dt))$ is controllable.

generalize these concepts beyond the focus on the state, which is a construct, and are merely one example of the system trajectory that one may want to steer or deduce from observations.

CONTROL AS INTERCONNECTION

System theory has two main historical roots, namely electrical circuit theory and automatic control. When system theory emerged as an independent discipline in the early 1960s, circuit theory was more developed and sophisticated than control. Nevertheless, it was under the impetus of control and filtering problems that system theory gained its momentum. In many ways, system theory can be viewed as an offshoot of control, while in theoretical developments, control maintains a central place. This origin may explain why input/output thinking, with its immediate relevance to sensor-output-to-actuator-input feedback control, acquired the status of the main paradigm for systems that interact with their environment.

Controlled systems—the interconnection of a plant with a controller—are examples of interconnected systems *par excellence*. We have learned to think of control in terms of inputs, outputs, and feedback. The resulting mechanism of feedback acting on a plant (see Figure 15), can be viewed as *intelligent control*, since the feedback mechanism

2) The rank of the complex matrix $R(\lambda)$ is the same for all $\lambda \in \mathcal{C}$.
3) There exists a polynomial matrix $M \in \mathbb{R}[\xi]^{\mathtt{w}\times\bullet}$ such that $\mathcal{B} = \mathtt{image}(M(d/dt))$.
4) The $\mathbb{R}[\xi]$-module spanned by the rows of R is closed (see "Polynomial Modules and Syzygies" for an explanation of the terminology).

Statement 3 in terms of an image is the most elegant test. For linear constant-coefficient differential operators, an image is always a kernel, while a kernel is an image only if the kernel defines a controllable behavior.

Statement 4 leads to a computable test. Closedness of the module requires that the syzygy of the right syzygy of R equals the module generated by the transposes of the rows of R. Verification of this condition leads to numerical test for controllability, since syzygies are computable using computer algebra. The image representation condition for controllability, Statement 3, as well as the closedness test, Statement 4, apply verbatim to linear constant-coefficient PDEs [34]. The existence of an image representation as a necessary and sufficient condition for controllability is also valid for delay-differential equations [35], [36].

Applied to $(d/dt)x = Ax + Bu$, $w = (x, u)$, where it is easy to see that state controllability is equivalent to behavioral controllability, Statement 2 yields the Popov-Belevich-Hautus test for state controllability, requiring the rank of $[A - I\lambda \; B]$ to be $\mathtt{dimension}(x)$ for all $\lambda \in \mathbb{C}$. It is well known that this condition in turn is equivalent to the condition

$$\mathtt{rank}\left([B \;\; AB \cdots A^{\mathtt{dimension}(x)-1}B]\right) = \mathtt{dimension}(x).$$

Applied to the single-input/single-output system

$$p\left(\frac{d}{dt}\right)y = q\left(\frac{d}{dt}\right)u, \quad w = (u, y)$$

with p and q real polynomials, the rank condition of Statement 2 shows that this differential equation defines a controllable system if and only if p and q do not have common factors. This result explains—at last—the notorious common factor problem, namely, that common factors in p and q correspond to a lack of controllability. Nothing more, nothing less.

That there is a connection between controllability and pole/zero cancellation has been an urban legend ever since the notion of state controllability was introduced. In one of Kalman's first articles on this subject, he states

> *A single input/single output plant (continuous-time or discrete-time) is completely controllable if and only if the input excites all natural frequencies of the plant; in other words, if no cancellation of poles is possible in the transfer function.* [S14, p. 484]

This statement is true, provided

i) we consider it as a statement about behavioral controllability—as such, this cancellation condition for controllability is a theorem *avant la lettre*
ii) we interpret the cancellation as referring to common roots of the polynomials p and q. The transfer function q/p is a rational function, and hence referring to common roots of its numerator and denominator doesn't make much sense mathematically, since common factors in the numerator and denominator of a rational function can, by definition, be canceled and introduced *ad libitum*.

With the behavioral definition of controllability, the common factor cancellation myth has been elevated to a fact, better yet, to a theorem. We refer to the classical notion that pertains to state-space systems, as *state controllability*, and to the behavioral notion, simply as *controllability*. Behavioral controllability makes controllability into an honest-to-goodness property of a dynamical system.

REFERENCES

[S11] R.E. Kalman, *Lectures on Controllability and Observability*, Centro Internazionale Matematico Estivo, Bologna, Italy, 1968.
[S12] C.A. Desoer, *Notes for a Second Course on Linear Systems*, New York: Van Nostrand Reinhold, 1970.
[S13] S. Shankar, "The evolution of the concept of controllability," *Math. Comput. Modelling Dynamical Syst.*, vol. 8, pp. 397–406, 2002.
[S14] R.E. Kalman, "On the general theory of control systems," *Automatic and Remote Control*, in *Proc. 1st Int. Congr. IFAC*, 1960, pp. 481–492.

is designed to act on the basis of observations, adapt to the environment, and make decisions. The observed sensor outputs determine the actuator inputs. In open-loop control, the environment acts as a signal generator and imposes an input signal on the system to be controlled. While it is not easy to define the exact nature of feedback [29], we take the presence of sensing and choosing control inputs on the basis of observations as the essential feature of feedback control.

However, many practical control devices do not act as feedback controllers. One example is passive suspension springs and dampers, which are used in automobiles to attenuate road irregularities and improve driving comfort (see Figure 16). These technological devices are control mechanisms, expressly designed to reduce the effect of disturbances in certain system variables, the acceleration of the body mass. But it makes little sense to view such control mechanisms as feedback controllers. The interaction of the suspension with the chassis and the axle of the car involves four variables, namely, the position of the lower terminal of the suspension and the force acting on it, as well as the position of the upper terminal of the suspension and the force acting on it. In viewing a suspension as a feedback controller, the question arises as to which of

these variables are to be considered sensed outputs and which variables are to be considered as control inputs. But this question arises only if one insists on viewing each system in an input/output way, and on viewing each controller in terms of sensors and actuators. The suspension is simply a subsystem embedded in the overall system to improve performance. The action of such passive controllers is best understood through interconnection and variable sharing, rather than signal processing and feedback control. The design of passive suspensions involves the synthesis of mechanical impedances [30], [31]. Numerous technological devices act and are designed as controllers, but are simply not feedback controllers. Examples include fins to radiate away excessive heat, insulation equipment for noise abatement, pressure valves, strips and grooves to suppress turbulence, and stabilizers for ships. Viewing control as subsystem design, with feedback control and open-loop control as special cases, greatly enhances the scope of the systems and control field.

From the behavioral point of view, control means restricting the behavior of a system, namely, the plant, through interconnection with another system, namely, the controller. This interconnection structure, shown in Figure 15, starts with a *plant*, a dynamical system with two sets of terminals. The variables on the first set of terminals are the *to-be-controlled variables* and belong to the set denoted by $\mathbb{V}$,

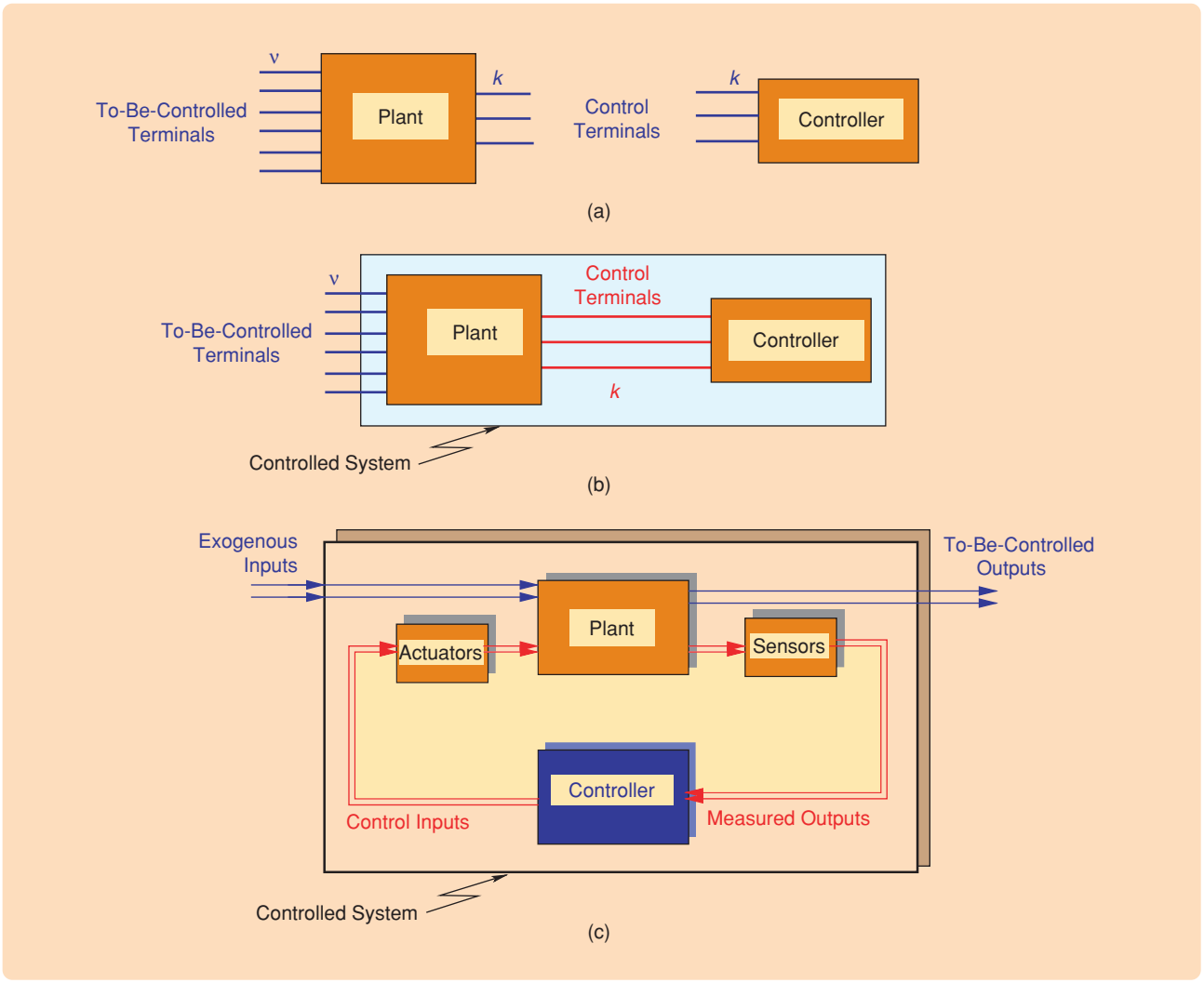


FIGURE 15 Control as interconnection. In behavioral control, the plant is viewed as a system with two types of terminals, a first set of terminals for the to-be-controlled variables and a second set for the control variables. The control variables are the variables that are available to the controller, such as sensor outputs and actuator inputs in intelligent control, or variables on the terminals where a controller can be interconnected to the plant in passive control. The to-be-controlled variables are variables such as disturbances, setpoints, and tracking signals that enter in the performance specifications of the control problem. The plant and controller are shown separately in (a). The controller is a system that imposes restrictions on the behavior of the control variables. After interconnection in (b), the plant and controller share the variables on the control terminals. This way, the restrictions imposed by the controller on the control variables are transmitted to the to-be-controlled variables of the plant. Classical sensor-output-to-actuator-input feedback (c) is a special case, with the control variables consisting of actuator inputs and sensor outputs shared by the plant and the controller.

while the variables on the second set of terminals are the *control variables* and belong to the set denoted by $\mathbb{K}$. The control variables are the variables that are available to the controller, such as sensor outputs and actuator inputs in feedback control, or variables on the terminals where a controller can be interconnected to the plant in passive control. The to-be-controlled variables are variables such as disturbances, setpoints, and tracking signals. The behavior of the to-be-controlled variables enters in the performance specifications of the control problem. The sets $\mathbb{V}$ and $\mathbb{K}$ need not be disjoint, since the to-be-controlled variables often include some of the control variables.

Note that we call all of the variables through which the plant interacts with the controller, control variables. In the case of a feedback controller these variables include the sensor outputs as well as the control inputs. Control variables should not be confused with control inputs.

The plant is thus a dynamical system

$$\Sigma_{\text{plant}} = (\mathbb{T}, \mathbb{V} \times \mathbb{K}, \mathcal{B}_{\text{plant}}),$$

where $\mathcal{B}_{\text{plant}} \subseteq (\mathbb{V} \times \mathcal{B})^{\mathbb{T}}$ denotes the *plant behavior*. The plant behavior relates the to-be-controlled variables $w : \mathbb{T} \to \mathbb{V}$ with the control variables $k : \mathbb{T} \to \mathbb{K}$. Without control, the plant allows the behavior of the to-be-controlled variables given by

$$\mathcal{B}_{\text{uncontrolled}} = \{v : \mathbb{T} \to \mathbb{V} \mid \text{there exists } k : \mathbb{T} \to \mathbb{K} \text{ such that } (v, k) \in \mathcal{B}_{\text{plant}}\},$$

obtained by eliminating the control variables from the plant behavior. The *controller* is a dynamical system

$$\Sigma_{\text{controller}} = (\mathbb{T}, \mathbb{K}, \mathcal{K}),$$

where $\mathcal{K} \subseteq \mathbb{K}^{\mathbb{T}}$ denotes the *controller behavior*. The controlled system is obtained by interconnecting the control terminals of the plant with the controller terminals. The interconnection restricts the control variables $k : \mathbb{T} \to \mathbb{K}$ of the plant to belong to $\mathcal{K}$. This way, the restrictions imposed on the control variables by the controller are transmitted to the to-be-controlled variables to meet the control specifications. Attaching the controller to the plant leads to a latent-variable representation of the full behavior of the *controlled system* given by

$$\mathcal{B}_{\text{full}} = \{(w, k) : \mathbb{T} \to \mathbb{K} \mid (w, k) \in \mathcal{B}_{\text{plant}} \text{ and } k \in \mathcal{K}\}.$$

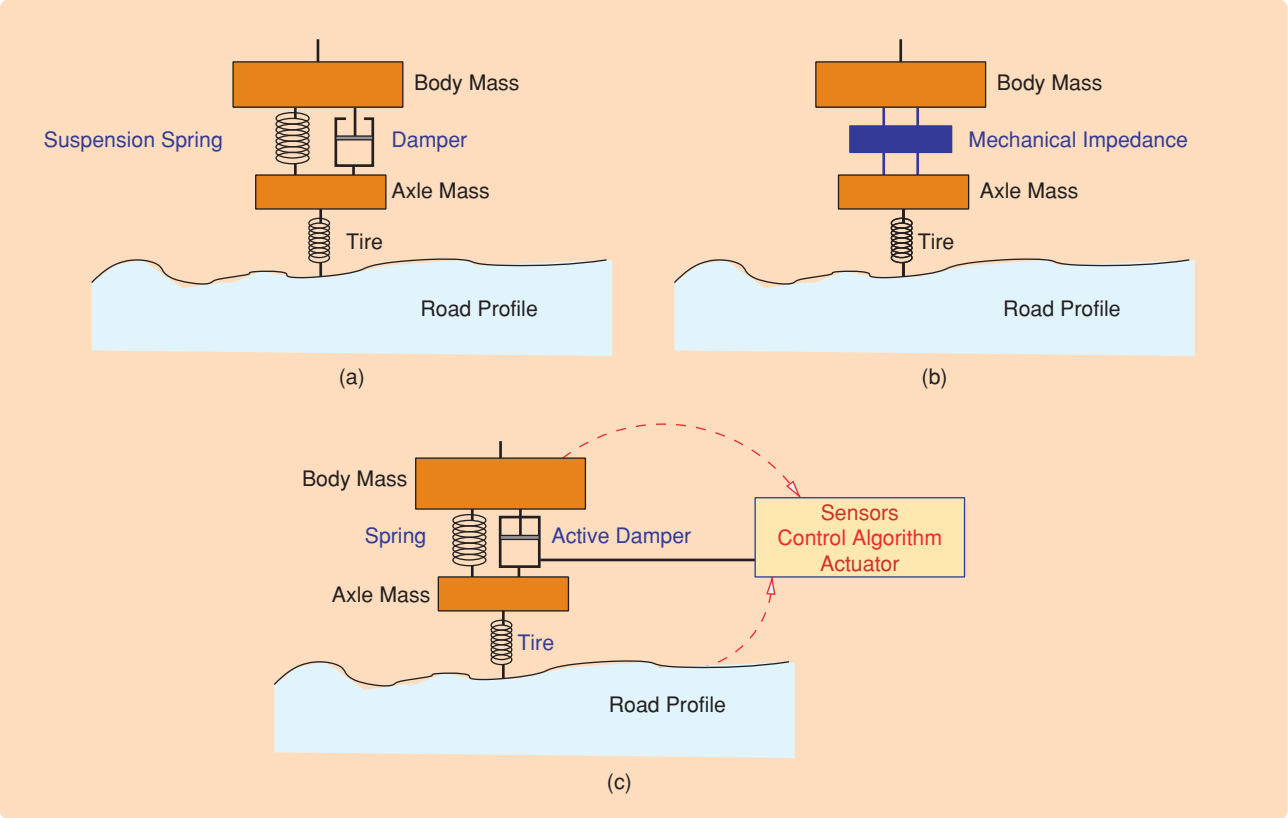


FIGURE 16 Passive and active suspensions. Part (a) illustrates a passive spring-damper suspension of a quarter-car model. This suspension can be viewed as a controller for suppressing the transmission of vibrations caused by the road. The controller is a mechanical impedance that interacts with the body and axle of the car by sharing position and forces at the interconnection points, as illustrated in (b). Despite common practice, it is unnatural to regard this controller as a feedback controller in terms of sensed outputs and actuated inputs. This passive controller can be contrasted with the active damper shown schematically in (c), which can be viewed as a feedback controller.

The *controlled system* $\Sigma_{\text{controlled}} = (\mathbb{T}, \mathbb{V}, \mathcal{B}_{\text{controlled}})$ has the behavior

$$\mathcal{B}_{\text{controlled}} = \{w : \mathbb{T} \to \mathbb{V} \mid \text{there exists } k \in \mathcal{K} \text{ such that } (w, k) \in \mathcal{B}_{\text{plant}}\},$$

obtained by eliminating the control variables from the full behavior. Obviously, $\mathcal{B}_{\text{controlled}} \subseteq \mathcal{B}_{\text{uncontrolled}}$, which shows that the controller restricts the behavior of the to-be-controlled variables in Σ_{plant}.

LINEAR TIME-INVARIANT SYSTEMS

The remaining sections of this article deal with linear time-invariant differential systems (LTIDSs), which are familiar through special cases, such as transfer functions, linear input/state/output models, as well as of differential-algebraic equations (DAEs). LTIDSs are prevalent for the following reasons.

1) A smooth nonlinear dynamical system allows a linear approximation, locally, around an operating point. In the time-invariant case, this approximation leads to an LTIDS. Linearization of a nonlinear behavior around an operating point is not a straightforward extension of the input/state/output case but can be defined appropriately.
2) Many physical laws, such as Newton's second law, Maxwell's equations, Kirchhoff laws, the Schrödinger equation of quantum mechanics, and the evolution of the probability density in stochastic differential equations, are linear and time invariant.
3) From a theoretical point of view, LTIDSs serve as a paradigm for more general and more challenging nonlinear systems.

Our aim is to explain some results that the behavioral approach brings for this familiar class of systems.

In contrast to the previous sections of this article, where motivation from physical systems dominates the discussion, this part is more mathematical in nature. LTIDSs and their mathematical description usually constitute one of the first topics covered in introductory systems and control courses. The approach pursued here is somewhat different from the usual one. By not starting with an input/output or input/state/output description, the modeling pitfalls discussed above are avoided. By taking the behavior as the central object of study, concepts such as controllability do not depend on a particular representation. Also, a clearer view of the role and limitations of transfer functions is obtained.

A few words about the notation used. $\mathcal{C}^\infty(\mathbb{R}\mathbb{R}^{\mathtt{w}})$ denotes the set of infinitely differentiable maps from the domain $\mathbb{R}$ to the codomain $\mathbb{R}^{\mathtt{w}}$, with obvious changes in the notation for other domains and codomains. $\mathbb{R}[\xi]$ denotes the set of polynomials, while $\mathbb{R}(\xi)$ denotes the set of rational functions with real coefficients in the indeterminate ξ (see "Polynomial Modules and Syzygies"). $\mathbb{R}[\xi]^{\mathtt{n}}$ denotes the set of $\mathtt{n}$-dimensional polynomial vectors, while $\mathbb{R}[\xi]^{\mathtt{n}_1 \times \mathtt{n}_2}$ denotes the set of polynomial matrices with $\mathtt{n}_1$ rows and $\mathtt{n}_2$ columns. When the dimension of a vector is immaterial, but finite, we use $^\bullet$, with similar notation for matrices. Of course, when we add, multiply, or equate vectors and matrices, we assume that the dimensions are compatible.

Definition of Linear Time-invariant Differential Behaviors

Let $\Sigma = (\mathbb{T}, \mathbb{W}, \mathcal{B})$ denote a dynamical system with time axis $\mathbb{T} = \mathbb{R}$ and signal space $\mathbb{W} = \mathbb{R}^\bullet$. The dynamical system $\Sigma = (\mathbb{R}, \mathbb{R}^\bullet, \mathcal{B})$ is linear if $\mathcal{B}$ is a linear space, that is, if $w_1, w_2 \in \mathcal{B}$ and $\alpha \in \mathbb{R}$ imply $w_1 + w_2 \in \mathcal{B}$ and $\alpha w_1 \in \mathcal{B}$, time invariant if $\sigma^t\mathcal{B} = \mathcal{B}$ for all $t \in \mathbb{R}$, and differential if its behavior is the solution set of a system of differential equations. A LTIDS (more precisely its behavior, but we make no distinction between a system and its behavior) is defined as $\Sigma = (\mathbb{R}, \mathbb{R}^\bullet, \mathcal{B})$, with $\mathcal{B}$ the set of solutions of a family of linear constant-coefficient differential equations. The behavior $\mathcal{B}$ of an LTIDS can thus be defined in terms of matrices $R_0, R_1, \ldots, R_{\mathtt{n}} \in \mathbb{R}^{\bullet\times\bullet}$ as the set of infinitely differentiable solutions of the system of differential equations

$$R_0 w + R_1 \frac{d}{dt} w + R_2 \frac{d^2}{dt^2} w + \cdots + R_{\mathtt{n}} \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}} w = 0.$$

This system of differential equations can be written compactly as

$$R\left(\frac{d}{dt}\right) w = 0, \tag{25}$$

where $R \in \mathbb{R}[\xi]^{\bullet\times\bullet}$ is the polynomial matrix

$$R(\xi) := R_0 \xi + R_1 \xi + R_2 \xi^2 + \cdots + R_{\mathtt{n}} \xi^{\mathtt{n}}.$$

Hence (25) defines the dynamical system $\Sigma = (\mathbb{R}, \mathbb{R}^\bullet, \mathcal{B})$ with

$$\mathcal{B} = \left\{ w \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet) \,\middle|\, R\left(\frac{d}{dt}\right) w = 0 \right\}.$$

Since some of the rows of R can have degree zero, (25) may represent a DAE. We denote this behavior by $\mathcal{B} = \mathtt{kernel}(R(d/dt))$, since $\mathcal{B}$ is the kernel of the linear constant-coefficient differential operator $R(d/dt) : \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet) \to \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet)$.

Denote the set of LTIDSs or their behaviors by $\mathcal{L}^\bullet$, and by $\mathcal{L}^{\mathtt{w}}$ when the number of variables, that is, the dimension of the vector w in (25), is $\mathtt{w}$. The classical single-input/single-output system

$$p_0 y + p_1 \frac{d}{dt} y + \cdots + p_{\mathtt{n}} \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}} y = q_0 u + q_1 \frac{d}{dt} u + \cdots + q_{\mathtt{n}} \frac{d^{\mathtt{n}}}{dt^{\mathtt{n}}} u,$$

written as (25), becomes

$$p\left(\frac{d}{dt}\right)y = q\left(\frac{d}{dt}\right)u$$

where

$$p(\xi) = p_0 + p_1\xi + \cdots + p_{\mathtt{n}}\xi^{\mathtt{n}}$$

and

$$q(\xi) = q_0 + q_1\xi + \cdots + q_{\mathtt{n}}\xi^{\mathtt{n}}.$$

Hence in this example $\mathtt{w} = 2$, $w = \begin{bmatrix} u \\ y \end{bmatrix}$, and $R = [\, q \quad -p\,]$.

Note that $\mathcal{L}^\bullet$ is defined in terms of the kernel representation (25), as consisting of those subsets $\mathcal{B}$ of $\mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet)$ such that $\mathcal{B} = \mathtt{kernel}(R(d/dt))$ for some polynomial matrix $R \in \mathbb{R}[\xi]^{\bullet\times\bullet}$. The analogous discrete-time system can be defined without invoking a representation. Indeed, in the discrete-time case, it can be shown that the following statements about a subset $\mathcal{B} \subseteq (\mathbb{R}^\bullet)^{\mathbb{Z}}$ are equivalent:

1) There exists $R \in \mathbb{R}[\xi]^{\bullet\times\bullet}$ such that $\mathcal{B} = \mathtt{kernel}(R(\sigma))$.
2) $\mathcal{B}$ is linear, shift invariant, and closed in the topology of pointwise convergence.
3) $\mathcal{B}$ is linear, shift invariant, and *complete*, that is, $w \in \mathcal{B}$ if and only if

$$w|_{[t_0,t_1]} \in \mathcal{B}|_{[t_0,t_1]} \quad \text{for all} \quad t_0, t_1 \in \mathbb{Z}. \tag{26}$$

Note that $w \in \mathcal{B}$ trivially implies (26), but the converse, that (26) implies $w \in \mathcal{B}$, requires structure on $\mathcal{B}$, and leads to the conclusion that the linear time-invariant behavior $\mathcal{B}$ is the set of solutions to a difference equation.

In the discrete-time case, the representation of $\mathcal{B}$ in terms of a difference equation, more precisely, as the kernel of a difference operator, can thus be deduced from high level, representation-free properties of the behavior. An analogue of Statement 3 can be obtained in the continuous-time case to define an LTIDS. It is, of course, desirable to define system classes and properties starting from intrinsic properties of the behavior, rather than in terms of representations.

It is interesting to connect $\mathcal{L}^\bullet$ to the classical input/output descriptions of linear systems in terms of transfer functions and input/state/output models. The connection can be briefly explained as follows. For every controllable system $\mathcal{B} \in \mathcal{L}^\bullet$, there exists a componentwise partition of the system variables w into $w \cong (u, y)$ ($\cong$ means that equality holds up to reordering of the components), such that $y = Gu$, where G is a matrix of proper rational functions, represents $\mathcal{B}$. In the section "Transfer Functions," it is explained what "represents" means here, in other words, how the behavior of $y = Gu$ is defined. It turns out that controllability is crucial here since only controllable systems can be described by a transfer function. On the other hand, for every $\mathcal{B} \in \mathcal{L}^\bullet$, there exists a componentwise partition of w into $w \cong (u, y)$, and polynomial matrices $P, Q \in \mathbb{R}[\xi]^{\bullet\times\bullet}$, with P square, $\det(P) \neq 0$, and $P^{-1}Q$ proper, such that $P(d/dt)y = Q(d/dt)u$ represents $\mathcal{B}$. Also, every $\mathcal{B} \in \mathcal{L}^\bullet$ admits a componentwise partition of the system variables w into $w \cong (u, y)$ and matrices A, B, C, D such that the representation $(d/dt)x = Ax + Bu$, $y = Cx + Du$ is a latent-variable representation of $\mathcal{B}$ (see "The Initial Value Problem"). The behavior of an LTIDS can always be described faithfully by a suitable state model, while transfer functions can exactly describe the behavior of only controllable systems. However, neither transfer functions nor state representations are good starting points for a theory of dynamics since they are special representations, and not the result of a modeling process.

The study of LTIDSs from a behavioral point of view motivates a wide range of theoretical questions [26], [32]. In the next subsections we concentrate on the following highlights, summarized in "Three Central Theorems":

1) the one-to-one correspondence between $\mathcal{L}^\bullet$ and $\mathbb{R}[\xi]$-modules.
2) the elimination theorem
3) controllability and the existence of an image representation.

Differential Annihilators

Let $w \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{w}})$. The polynomial vector $n \in \mathbb{R}[\xi]^{\mathtt{w}}$ is a *linear differential annihilator*, or simply, an *annihilator* of w if $n(d/dt)^\top w = 0$. This definition is readily extended from $w \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{w}})$ to a behavior $\mathcal{B} \subseteq \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{w}})$, by requiring $n(d/dt)^\top \mathcal{B} = 0$, that is, $n(d/dt)^\top w = 0$ for all $w \in \mathcal{B}$. Let $\mathcal{N}_{\mathcal{B}}^{\mathbb{R}[\xi]} \subseteq \mathbb{R}[\xi]^{\mathtt{w}}$ denote the set of linear differential annihilators of $\mathcal{B}$. Note that when $\mathcal{B} \in \mathcal{L}^{\mathtt{w}}$ has (25) as a kernel representation, then $\mathcal{B} = \mathtt{kernel}(R(d/dt))$ consists of those $w \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{w}})$ that have the transposes of the rows of R as annihilators. When $\mathcal{B} \subseteq \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{w}})$ is shift invariant and closed, $\mathcal{N}_{\mathcal{B}}^{\mathbb{R}[\xi]} \subseteq \mathbb{R}[\xi]^{\mathtt{w}}$ has a nice mathematical structure, specifically, $\mathcal{N}_{\mathcal{B}}^{\mathbb{R}[\xi]} \subseteq \mathbb{R}[\xi]^{\mathtt{w}}$ is an $\mathbb{R}[\xi]$-module (see "Polynomial Modules and Syzygies" for this terminology). This module structure is easy to verify. The sum of two annihilators is again an annihilator, and, if we premultiply an annihilator by a polynomial, we again obtain an annihilator. Since $\mathbb{R}[\xi]^{\mathtt{w}}$ is itself also an $\mathbb{R}[\xi]$-module, $\mathcal{N}_{\mathcal{B}}^{\mathbb{R}[\xi]}$ is a submodule of the $\mathbb{R}[\xi]$-module $\mathbb{R}[\xi]^{\mathtt{w}}$. Henceforth, when $\mathcal{N}_{\mathcal{B}}^{\mathbb{R}[\xi]}$ or $\mathbb{R}[\xi]^{\mathtt{w}}$ is referred to as a module, it is understood that this nomenclature means $\mathbb{R}[\xi]$-module.

When $\mathcal{B} = \mathtt{kernel}\,(R(d/dt))$, each n of the form $n = (fR)^\top$ for some $f \in \mathbb{R}[\xi]^{1\times\bullet}$ is an annihilator. Hence the module of annihilators $\mathcal{N}_{\mathcal{B}}^{\mathbb{R}[\xi]}$ contains the module generated by the transposes of the rows of R. The question arises as to whether these two modules are equal, that is, whether every annihilator of $\mathcal{B}$ is actually a linear combination with polynomial coefficients of the rows of the polynomial matrix R of the system $R(d/dt)w = 0$ of differential equations that defines $\mathcal{B}$.

To appreciate what is involved in the equality of these two modules, consider, as an example, the scalar differential equation $p(d/dt)w = 0$ with $0 \neq p \in \mathbb{R}[\xi]$. Let $q \in \mathbb{R}[\xi]$. It is easy to prove that

Polynomial Modules and Syzygies

When doing linear system theory from a behavioral point of view, it is easy to become enamored with polynomial modules. Here we try to explain the object of this infatuation.

Polynomials are best viewed in terms of indeterminates. In other words, in the expression $\pi(\xi) = \pi_0 + \pi_1\xi + \cdots + \pi_{\mathtt{m}}\xi^{\mathtt{m}}$, with the coefficients $\pi_{\mathtt{k}}$ real numbers, ξ is viewed as an abstract variable, called an *indeterminate*. This interpretation allows us to substitute various mathematical objects for ξ, for example a real number, leading to the map $x \in \mathbb{R} \mapsto \pi(x) \in \mathbb{R}$, or a complex number, leading to the map $s \in \mathbb{C} \mapsto \pi(s) \in \mathbb{C}$, or a square matrix, leading to the map $A \in \mathbb{R}^{\mathtt{n}\times\mathtt{n}} \mapsto \pi(A) \in \mathbb{R}^{\mathtt{n}\times\mathtt{n}}$, or d/dt, leading to the differential operator $\pi(d/dt)$. The set of polynomials with real coefficients is denoted as $\mathbb{R}[\xi]$. $\mathbb{R}[\xi]$ is a ring, meaning that elements can be added and multiplied, and that these operations satisfy a number of requirements that are evident.

The set of $\mathtt{w}$-dimensional polynomial vectors $\mathbb{R}[\xi]^{\mathtt{w}}$ has the mathematical structure of a *module*. A module $\mathcal{M}$ over a ring $\mathcal{R}$ comes equipped with a commutative binary operation that allows elements of $\mathcal{M}$ to be added, and with scalar multiplication that allows elements of $\mathcal{M}$ to be multiplied by scalars from $\mathcal{R}$. Hence, for $m_1, m_2 \in \mathcal{M}$ and $r \in \mathcal{R}$, $m_1 + m_2 \in \mathcal{M}$ and $rm_1 \in \mathcal{M}$. Of course, these operations need to satisfy certain properties, which are usually evident. $\mathcal{M}$ is called an *$\mathcal{R}$-module*, unless the ring $\mathcal{R}$ is obvious from the context. Informally, one can think of a module as a vector space in which the scalars are elements of a ring, rather than of a field, as required for vector spaces. Working with modules is like doing linear algebra over a ring, and is often more fun than over a field. $\mathbb{R}[\xi]^{\mathtt{w}}$ is a module over the ring $\mathbb{R}[\xi]$, since elements of $\mathbb{R}[\xi]^{\mathtt{w}}$ can be added and multiplied by scalars from the ring $\mathbb{R}[\xi]$.

A module $\mathcal{M}$ is *finitely generated* if there exist *generators*, $g_1, g_2, \ldots, g_{\mathtt{p}} \in \mathcal{M}$ such that, for each element $m \in \mathcal{M}$, there exist $\alpha_1, \ldots, \alpha_{\mathtt{p}} \in \mathcal{R}$ such that $m = \alpha_1 g_1 + \cdots + \alpha_{\mathtt{p}} g_{\mathtt{p}}$. If the generators can be chosen to be *independent*, meaning that

$$\alpha_1 g_1 + \cdots + \alpha_{\mathtt{p}} g_{\mathtt{p}} = 0$$

implies

$$\alpha_1 = \cdots = \alpha_{\mathtt{p}} = 0,$$

then $\mathcal{M}$ is *free*. A set of independent generators of a free module $\mathcal{M}$ is a *basis* for $\mathcal{M}$, while the number of elements of a basis is the *dimension* of $\mathcal{M}$, denoted *dimension* $(\mathcal{M})$. A subset of an $\mathcal{R}$-module $\mathcal{M}$ that is closed under addition and under scalar multiplication by elements from $\mathcal{R}$ is an $\mathcal{R}$- *submodule* of $\mathcal{M}$. All $\mathbb{R}[\xi]$-submodules of $\mathbb{R}[\xi]^{\mathtt{w}}$ (here $\mathcal{R} = \mathbb{R}[\xi]$) are finitely generated and free and thus we can speak of the dimension of such a submodule. Below, we meet $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$, the polynomials in many variables. Every $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$-submodule of $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]^{\mathtt{w}}$ is finitely generated but may not be free.

Submodules of $\mathbb{R}[\xi]^{\mathtt{w}}$ are useful in system theory since there is a one-to-one correspondence between linear time-invariant differential systems with $\mathtt{w}$ variables, and the $\mathbb{R}[\xi]$-submodules of $\mathbb{R}[\xi]^{\mathtt{w}}$. This correspondence associates with an LTIDS $\mathcal{B}$ the module of its differential annihilators, that is, the polynomial vectors $n \in \mathbb{R}[\xi]^{\mathtt{w}}$ such that $n(d/dt)^{\top}\mathcal{B} = 0$. The set of differential annihilators of the LTIDS defined as the $\mathcal{C}^{\infty}(\mathbb{R}, \mathbb{R}^{\mathtt{w}})$-solutions of $R(d/dt)w = 0$, where $R \in \mathbb{R}[\xi]^{\bullet\times\mathtt{w}}$, is equal to the $\mathbb{R}[\xi]$-module generated by the transposes of the rows of R. This result depends on the fact that $\mathcal{C}^{\infty}$-solutions are used.

The one-to-one correspondence between LTIDSs and $\mathbb{R}[\xi]$-submodules of $\mathbb{R}[\xi]^{\mathtt{w}}$ implies that $R_1(d/dt)w = 0$ and $R_2(d/dt)w = 0$ have the same behavior if and only if the rows of R_1 and R_2 generate the same $\mathbb{R}[\xi]$-module. In other words, if and only if each row of R_1 is a linear combination with polynomial coefficients of the rows of R_2, and vice-versa, each row of R_1 is a linear combination with polynomial coefficients of the rows of R_2. Hence, if these rows form a basis for the corresponding module, equivalently, if the polynomial matrices R_1 and R_2 are of full row rank, then the behavior defined by $R_1(d/dt)w = 0$ is equal to the behavior defined by $R_2(d/dt)w = 0$ if and only if there exists a unimodular polynomial matrix U such that $R_1 = UR_1$. The one-to-one correspondence fails if we use solutions with compact support, but remains valid for distributional solutions. The concordance between LTIDSs and submodules of $\mathbb{R}[\xi]^{\mathtt{w}}$ gives submodules a prominent place in the field. Linear systems theory deals with submodules. Every property of an LTIDS can be translated into a property of the corresponding submodule, and vice versa. We illustrate this correspondence by means of controllability.

The *closure* of an $\mathbb{R}[\xi]$-submodule $\mathcal{M}$ of $\mathbb{R}[\xi]^{\mathtt{w}}$ is defined as

$$\bar{\mathcal{M}} := \{\bar{m} \in \mathbb{R}[\xi]^{\mathtt{w}} \mid \text{there exist } \pi \in \mathbb{R}[\xi], \quad \pi \neq 0, \text{ and } m \in \mathcal{M} \text{ such that } m = \pi\bar{m}\}.$$

$\bar{\mathcal{M}}$ is an $\mathbb{R}[\xi]$-submodule of $\mathbb{R}[\xi]^{\mathtt{w}}$. If $\mathcal{M} = \bar{\mathcal{M}}$, then $\mathcal{M}$ is *closed*. A submodule $\mathcal{M}$ of $\mathbb{R}[\xi]^{\mathtt{w}}$ is closed if and only if $\mathcal{M}$ is not properly contained in any $\mathbb{R}[\xi]$-submodule of $\mathbb{R}[\xi]^{\mathtt{w}}$ of the same dimension.

It can be shown that the behavior $\mathcal{B}$ is controllable (in the behavioral sense, of course) if and only if the corresponding submodule of annihilators is closed. More generally, if the behavior $\mathcal{B}$ corresponds to the submodule $\mathcal{M}$, then $\bar{\mathcal{M}}$ corresponds to the controllable part of $\mathcal{B}$.

Consider, for an $\mathbb{R}[\xi]$-submodule $\mathcal{M}$ of $\mathbb{R}[\xi]^{\mathtt{w}}$, the set

$$\mathcal{M}^{\perp} := \{n \in \mathbb{R}[\xi]^{\mathtt{w}} \mid n^{\top}m = 0 \quad \text{for all} \quad m \in \mathcal{M}\},$$

which is the *syzygy* of $\mathcal{M}$. $\mathcal{M}^{\perp}$ is also an $\mathbb{R}[\xi]$-submodule of $\mathbb{R}[\xi]^{\mathtt{w}}$. It is easy to see that $\mathcal{M}^{\perp}$ is closed and satisfies $(\mathcal{M}^{\perp})^{\perp} = \bar{\mathcal{M}}$, $\mathcal{M} \cap \mathcal{M}^{\perp} = \{0\}$, and $\mathtt{dimension}\ (\mathcal{M}) + \mathtt{dimension}\ (\mathcal{M}^{\perp}) = \mathtt{w}$. However, contrary to what is the case for subspaces of $\mathbb{R}^{\mathtt{w}}$, $\mathcal{M}$ and $\mathcal{M}^{\perp}$ may not be direct summands and $\mathcal{M} + \mathcal{M}^{\perp}$ may be a proper subset of $\mathbb{R}[\xi]^{\mathtt{w}}$. In fact, the following conditions are equivalent:

1) $\mathcal{M}$ is closed.

2) $\mathcal{M}$ has a direct summand.

3) $\mathcal{M} \oplus \mathcal{M}^{\perp} = \mathbb{R}[\xi]^{\mathtt{w}}$.

4) $\mathcal{M} = (\mathcal{M}^{\perp})^{\perp}$.

A simple example to illustrate that submodules behave differently from subspaces is provided by the case $\mathtt{w} = 1$. The submodules of $\mathbb{R}[\xi]$ are precisely the subsets of the form $\mathbb{R}[\xi]\pi$, that is, the polynomials that have π as a factor, with $\pi \in \mathbb{R}[\xi]$. For $\pi \neq 0$, each of these submodules has dimension 1, its syzygy equals $\{0\}$, and its closure is all of $\mathbb{R}[\xi]$. Hence $\mathcal{M} + \mathcal{M}^{\perp} = \mathbb{R}[\xi]$ if and only if either $\pi = 0$ or $\mathtt{degree}(\pi) = 0$. In fact, $\mathbb{R}[\xi]$ and $\{0\}$ are the only submodules that are not properly contained in a submodule of the same dimension.

We use the *syzygy* terminology also for a polynomial matrix $F \in \mathbb{R}[\xi]^{\mathtt{w}_1 \times \mathtt{w}_2}$, as follows. Obviously, the sets

$$\{n_1 \in \mathbb{R}[\xi]^{\mathtt{w}_1} \mid n_1^{\top} F = 0\}$$

and

$$\{n_2 \in \mathbb{R}[\xi]^{\mathtt{w}_2} \mid F n_2 = 0\}$$

are $\mathbb{R}[\xi]$-modules. They are equal to the syzygies of the modules generated by the columns and transposes of the rows of F, respectively. We refer to these submodules as the *left syzygy of* F and the *right syzygy of* F, respectively. It follows that $R(d/dt)w = 0$ is controllable if and only if the syzygy of the right syzygy of R is equal to the module generated by the transposes of the rows of R.

We now look at the generalization to multivariable polynomials. The ring of polynomials in $\mathtt{n}$ indeterminates with real coefficients is denoted by $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$. $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]^{\mathtt{w}}$, the set of $\mathtt{w}$-dimensional vectors whose components are elements of $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$, is an $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$-module. As in the case $\mathtt{n} = 1$, its submodules are finitely generated, but in the case $\mathtt{n} > 1$, they may not be free, contrary to the case $\mathtt{n} = 1$. For example, the submodule $\xi_1\mathbb{R}[\xi_1, \xi_2] + \xi_2\mathbb{R}[\xi_1, \xi_2]$ of $\mathbb{R}[\xi_1, \xi_2]$ is not free. The notion of syzygy carries over, unchanged.

One of the nice aspects of the behavioral approach is that the theory of dynamical systems with one independent variable, time, carries over, with much the same concepts and similar notation to spatially distributed systems with many independent variables, usually space, or time and space. Monovariable polynomials are replaced by multivariable ones. This way, linear system theory extends smoothly to PDEs. Although most laws in physics are expressed in terms of PDEs, the study of systems described by PDEs has not achieved a central role in systems and control theory. Introductory systems courses hardly mention PDEs.

Let $R \in \mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]^{\bullet \times \mathtt{w}}$ and consider the system of linear constant-coefficient PDEs

$$R\left(\frac{\partial}{\partial v_1}, \frac{\partial}{\partial v_2}, \ldots, \frac{\partial}{\partial v_{\mathtt{n}}}\right) w = 0.$$

Its $\mathcal{C}^{\infty}(\mathbb{R}, \mathbb{R}^{\mathtt{w}})$-solutions form a behavior $\mathcal{B}$. The notion of annihilator carries over and so does the question of whether the $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$ submodule formed by the annihilators is equal to the $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$ submodule generated by the rows of R. Equivalently, there is a one-to-one relation between the linear shift-invariant differential $\mathtt{n}$D behaviors and the $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$ submodules of $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]^{\mathtt{w}}$. In [43] it is shown that this one-to-one relation holds, by proving that $\mathcal{C}^{\infty}(\mathbb{R}, \mathbb{R})$, viewed as a $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$-module, is an "injective cogenerator." Hence, as in the case $\mathtt{n} = 1$, $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$ submodules of $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]^{\mathtt{w}}$ stand in one-to-one correspondence with $\mathcal{C}^{\infty}$-behaviors of linear constant-coefficient PDEs. Consequently, every property of the corresponding $\mathtt{n}$D system can again be deduced from the associated module. The relation between submodules and linear $\mathtt{n}$D systems is discussed in [S15]–[S19].

As in the case $\mathtt{n} = 1$, the linear constant-coefficient PDE defines a controllable system if and only if the submodule of annihilators is closed, equivalently, if and only if the syzygy of the right syzygy of R is equal to the module generated by the transposes of the rows of R. Tests for controllability therefore involve the computation of syzygies. Algorithms for computing syzygies and similar objects are available in computer algebra packages [12].

This algebraic approach to linear system theory extends to other classes, such as discrete-time systems with the ring $\mathbb{R}[\xi]$ or $\mathbb{R}[\xi, \xi^{-1}]$, depending on whether the time axis is $\mathbb{Z}_+$ or $\mathbb{Z}$, to 1D- and $\mathtt{n}$D-convolutional codes, where it is customary to consider compact support signals or fields, to delay-differential systems with the ring $\mathbb{R}[\xi_1, \xi_2, \xi_2^{-1}]$, and to $\mathtt{n}$D difference equations with the ring $\mathbb{R}[\xi_1, \xi_1^{-1}, \xi_2, \xi_2^{-1}, \ldots, \xi_{\mathtt{n}}, \xi_{\mathtt{n}}^{-1}]$. Extensions to time-varying systems are also possible [S20], [S21], but are more difficult to obtain.

REFERENCES

[S15] E. Zerz, "Extension modules in behavioral linear systems theory," *Multidimensional Syst. Signal Processing*, vol. 12, pp. 309–327, 2001.

[S16] J. Wood, "Modules and behaviors in $\mathtt{n}$D systems theory," *Multidimensional Syst. Signal Processing*, vol. 11, pp. 11–48, 2000.

[S17] J. Wood, "Key problems in the extension of module-behavior duality," *Linear Algebra Its Applications*, vol. 351–352, pp. 761–798, 2002.

[S18] S. Shankar, "The Nullstellensatz for systems of PDE," *Adv. Appl. Math.*, vol. 23, pp. 360–374, 1999.

[S19] A.S. Sasane, "On the Willems closure with respect to W_S," *IMA J. Math. Contr. Inform.*, vol. 20, pp. 217–232, 2003.

[S20] A. Ilchmann and V. Mehrmann, "A behavioral approach to time-varying systems," *SIAM J. Control Optim.*, vol. 44, pp. 1725–1747, 2005.

[S21] E. Zerz, "An algebraic approach to time-varying systems," *IMA J. Math. Contr. Inform.*, vol. 23, pp. 113–126, 2006.

$$w \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}) \quad \text{and} \quad p\left(\frac{d}{dt}\right) w = 0$$

imply

$$q\left(\frac{d}{dt}\right) w = 0$$

if and only if p is a factor of q, that is, if and only if q belongs to the module generated by p. In this case, the two modules are equal. However,

$$w \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}) \text{ of compact support and } p\left(\frac{d}{dt}\right) w = 0$$

imply

$$w = 0.$$

Therefore, if we use compact support solutions, then every $q \in \mathbb{R}[\xi]$, not just those in the module generated by p, leads to an annihilator. Hence, if compact support solutions are considered, these polynomial modules are not equal.

We associate with each submodule $\mathcal{M}$ of $\mathbb{R}[\xi]^{\mathtt{w}}$ an element of $\mathcal{L}^{\mathtt{w}}$, as follows. Define

$$\mathcal{B}^{\mathcal{M}} := \left\{ w \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{w}}) \;\middle|\; m\left(\frac{d}{dt}\right)^\top w = 0, \text{ for all } m \in \mathcal{M} \right\}.$$

Obviously, $\mathcal{B}^{\mathcal{M}} = \mathtt{kernel}\,(G(d/dt))$, with $G \in \mathbb{R}[\xi]^{\bullet \times \mathtt{w}}$ a matrix whose rows are the transposes of a set of generators of $\mathcal{M}$. Hence $\mathcal{B}^{\mathcal{M}} \in \mathcal{L}^{\mathtt{w}}$. Thus each $\mathcal{B} \in \mathcal{L}^{\mathtt{w}}$ generates, through its set of linear differential annihilators, a submodule $\mathcal{N}_{\mathcal{B}}^{\mathbb{R}[\xi]}$ of $\mathbb{R}[\xi]^{\mathtt{w}}$, while, conversely, each $\mathbb{R}[\xi]$-submodule $\mathcal{M}$ of $\mathbb{R}[\xi]^{\mathtt{w}}$ generates a behavior $\mathcal{B}^{\mathcal{M}} \in \mathcal{L}^{\mathtt{w}}$. It turns out that this correspondence is bijective.

Theorem 1 (Submodule Theorem)

Let $\mathcal{B}$ be the behavior of (25). Then $\mathcal{N}_{\mathcal{B}}^{\mathbb{R}[\xi]}$ is equal to the $\mathbb{R}[\xi]$-module generated by the transposes of the rows of R. Furthermore, there exists a bijective correspondence between $\mathcal{L}^{\mathtt{w}}$ and the submodules of $\mathbb{R}[\xi]^{\mathtt{w}}$, the correspondence being

Three Central Theorems

A linear time-invariant differential system (LTIDS) has a behavior that is the set of solutions of a system of linear constant-coefficient differential equations

$$R\left(\frac{d}{dt}\right) \mathtt{w} = 0, \tag{S14}$$

with $R \in \mathbb{R}[\xi]^{\bullet \times \bullet}$ a polynomial matrix with real coefficients.The column dimension of R corresponds to the number of system variables, while the row dimension equals the number of differential equations that specify the behavior. Finite dimensional linear time-invariant state-space systems, transfer functions, and linear differential-algebraic equations are special cases. Infinitely differentiable ($\mathcal{C}^\infty$) solutions are usually considered. Alternative function spaces are also of interest but may influence the results. The theory of LTIDSs centers around linear constant-coefficient differential operators

$$P\left(\frac{d}{dt}\right) : \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet) \to \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet),$$

with $P \in \mathbb{R}[\xi]^{\bullet \times \bullet}$. An LTIDS has a behavior that is the kernel of such an operator.

Three central theorems give this class of systems a solid mathematical underpinning.

Fact 1: There is a one-to-one relation between LTIDSs with $\mathtt{w}$ variables and $\mathbb{R}[\xi]$-submodules of $\mathbb{R}[\xi]^{\mathtt{w}}$. This one-to-one relation associates with $\mathtt{kernel}(R(d/dt))$ the $\mathbb{R}[\xi]$-module generated by the rows of R. This result implies that each LTIDS allows a representation (S14) with R of full row rank, called a *minimal* kernel representation. Two minimal kernel representations $R_1(d/dt)w = 0$ and $R_2(d/dt)w = 0$ define the same system if and only if there exists a unimodular polynomial matrix U such that $R_1 = UR_2$.

Fact 2: Every image is a kernel. More precisely, for each $M \in \mathbb{R}[\xi]^{\mathtt{w} \times \bullet}$, there exists $R \in \mathbb{R}[\xi]^{\bullet \times \mathtt{w}}$ such that

$$\mathtt{kernel}\left(R\left(\frac{d}{dt}\right)\right) = \mathtt{image}\left(M\left(\frac{d}{dt}\right)\right).$$

This result implies the elimination theorem, a result that is useful in model building. The elimination theorem states that latent variables in LTIDSs can be completely eliminated, leading to a system of differential equations that describes the manifest behavior and contains only the manifest variables. The elimination theorem implies that the set of LTIDSs is closed under addition, intersection, and projection.

Fact 3: A kernel is an image if and only if the corresponding behavior is controllable. More precisely, for a given $R \in \mathbb{R}[\xi]^{\bullet \times \mathtt{w}}$, there exists $M \in \mathbb{R}[\xi]^{\mathtt{w} \times \bullet}$ such that

$$\mathtt{image}\left(M\left(\frac{d}{dt}\right)\right) = \mathtt{kernel}\left(R\left(\frac{d}{dt}\right)\right)$$

if and only if

$$\mathcal{B} = \mathtt{kernel}\left(R\left(\frac{d}{dt}\right)\right) \text{ is controllable.}$$

Mutatis mutandis, these results hold for discrete-time systems; this generalization is easy. But these three theorems also hold for systems described by linear constant-coefficient PDEs. This generalization is mathematically deep and relevant in bringing linear constant-coefficient PDEs into the realm of linear system theory.

$$\mathcal{B} \mapsto \mathcal{N}_{\mathcal{B}}^{\mathbb{R}[\xi]}, \quad \mathcal{M} \mapsto \mathcal{B}^{\mathcal{M}}.$$

In view of the preamble, the only thing that needs to be shown is that if $\mathcal{M}_1$ and $\mathcal{M}_2$ are distinct submodules of $\mathbb{R}[\xi]^{\mathbb{W}}$, then $\mathcal{B}^{\mathcal{M}_1}$ and $\mathcal{B}^{\mathcal{M}_2}$ are distinct elements of $\mathcal{L}^{\mathbb{W}}$. This property is easy to establish but is not trivial. In fact, as the example $p(d/dt)w = 0$ with $0 \neq p \in \mathbb{R}[\xi]$ shows, the result depends on the fact that we use $\mathcal{C}^\infty$-solutions. The submodule theorem is not valid if $\mathcal{C}^\infty$-solutions with compact support are considered to define the behavior.

The submodule theorem implies, in particular, that every $\mathcal{B} \in \mathcal{L}^{\mathbb{W}}$ satisfies $\mathcal{B} = \mathcal{B}^{\mathcal{N}_{\mathcal{B}}^{\mathbb{R}[\xi]}}$ and that every submodule $\mathcal{M}$ of $\mathbb{R}[\xi]^{\mathbb{W}}$ satisfies $\mathcal{M} = \mathcal{N}_{\mathcal{B}^{\mathcal{M}}}^{\mathbb{R}[\xi]}$. A behavior $\mathcal{B} \in \mathcal{L}^{\mathbb{W}}$ can therefore be thought of as the set of solutions of a module, hence an infinite set of linear constant-coefficient differential equations rather than a finite set, as (25) suggests.

The following statements are immediate consequences of the submodule theorem:

1) Let $\mathcal{B}_1$ be the behavior of $R_1(d/dt)w = 0$, and $\mathcal{B}_2$ be the behavior of $R_2(d/dt)w = 0$. Then $\mathcal{B}_1 \subseteq \mathcal{B}_2$ if and only if there exists $F \in \mathbb{R}[\xi]^{\bullet\times\bullet}$ such that $R_2 = FR_1$.
2) For all $\mathcal{B} \in \mathcal{L}^{\bullet}$, there exists a polynomial matrix $R \in \mathbb{R}[\xi]^{\bullet\times\bullet}$ of full row rank, meaning that the rank of R is equal to the number of row of R, such that $\mathcal{B} = \mathtt{kernel}(R(d/dt))$. A kernel representation (25) of $\mathcal{B}$ with R of full row rank is *minimal*, since the number of rows of R is then as small as possible among all kernel representations of $\mathcal{B}$.
3) $R_1(d/dt)w = 0$ and $R_2(d/dt)w = 0$ are minimal kernel representations of the same $\mathcal{B} \in \mathcal{L}^{\mathbb{W}}$ if and only if $R_2 = UR_1$ for some unimodular polynomial matrix U. A polynomial matrix is *unimodular* if it has an inverse that is also a polynomial matrix.

We mention the following application of Statement 3. Call the system $\mathcal{B} \in \mathcal{L}^{\mathbb{W}}$ *time reversible* if $w \in \mathcal{B}$ implies $\mathtt{rev}(w) \in \mathcal{B}$, where $\mathtt{rev}$ denotes *time reversal*, defined by $\mathtt{rev}(w) : t \mapsto w(-t)$. Consider the scalar system described by $p(d/dt)w = 0$ with $0 \neq p \in \mathbb{R}[\xi]$. Clearly, $p(-(d/dt))w = 0$ is a minimal kernel representation of $\mathtt{rev}(\mathcal{B})$. Therefore, $p(d/dt)w = 0$ describes a time-reversible system if and only if there exists $0 \neq \alpha \in \mathbb{R}$ such that $p(\xi) = \alpha p(-\xi)$. Consequently, time reversibility implies that p is either an even or an odd polynomial. A multivariable generalization of this result is given in [33].

The Elimination Theorem for LTIDSs

The elimination of auxiliary variables in modeling is motivated in the section "Latent Variables." In the context of LTIDSs, we thus have the following elimination question. If the behavior of a set of variables is an LTIDS, is a subset of these variables also an LTIDS? The elimination theorem states that for LTIDSs complete elimination of latent variables is possible.

Theorem 2. (Elimination Theorem)
Consider a behavior $\mathcal{B} \in \mathcal{L}^{\mathbb{W}_1+\mathbb{W}_2}$. Define

$$\mathcal{B}_1 = \{w_1 \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathbb{W}_1}) \mid \text{there exists } w_2 \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathbb{W}_2}) \text{ such that } (w_1, w_2) \in \mathcal{B}\}.$$

Then $\mathcal{B}_1 \in \mathcal{L}^{\mathbb{W}_1}$.

In "The Fundamental Principle and the Elimination Theorem," a proof of this theorem is outlined that extends to PDEs and can be implemented as an elimination algorithm for elimination of latent variables starting from a kernel representation.

Consider a system of linear constant-coefficient differential equations with both manifest variables w and latent variables ℓ, given by

$$R\left(\frac{d}{dt}\right)w = M\left(\frac{d}{dt}\right)\ell, \tag{27}$$

with $R, M \in \mathbb{R}[\xi]^{\bullet\times\bullet}$. We view (27) as describing the behavior of manifest variables w in terms of differential equations that also involve latent variables ℓ. Such models with latent variables are often the result of first principles modeling. The elimination theorem allows us to conclude that the manifest behavior is an LTIDS. In other words, there exists a polynomial matrix $R' \in \mathbb{R}[\xi]^{\bullet\times\mathbb{W}}$ such that the trajectories $w : \mathbb{R} \to \mathbb{R}^{\mathbb{W}}$ that (27) declares possible are precisely those that satisfy

$$R'\left(\frac{d}{dt}\right)w = 0. \tag{28}$$

To motivate the relevance of the elimination theorem in modeling, consider the following questions:

1) Does the w-behavior of the DAE $(d/dt)Ex = Ax + Bw$ described by a differential equation involve only w?
2) Consider an electrical circuit possessing external terminals and composed of an interconnection of linear resistors, capacitors, inductors, transformers, and gyrators. Is the behavior of the external voltage/current vector governed by a differential equation involving only the external voltages and currents?
3) Consider a mechanical system composed of an interconnection of linear springs, dampers, and masses. Is the terminal force/position behavior described by a differential equation that contains only the external forces and positions?

The answer to these questions is in the affirmative, and this conclusion is an immediate consequence of the elimination theorem. As long as the model equations are linear ordinary differential equations (ODEs) or DAEs with constant coefficients, the internal latent variables can be eliminated from the equations, and we end up again with a system of linear constant-coefficient ODEs involving only the external manifest variables.

It is somewhat surprising that this result does not play a more prominent role in system theory. Usually the degree of the polynomial matrix in a kernel representation of the manifest behavior (28) after elimination is higher than the degrees of the polynomial matrices in the latent-variable model (27) before elimination. However, regardless of the number of equations and the number of latent variables in (27), the system (28) of ODEs that describes the manifest behavior after elimination need never consist of more differential equations than the number of manifest variables. This result is a

The Fundamental Principle and the Elimination Theorem

The fundamental principle and the elimination theorem deal with both PDEs and ODEs. We state the fundamental principle for PDEs and the elimination theorem for ODEs.

Let $F \in \mathbb{R}[\xi_1, \xi_2, \ldots, \xi_n]^{y \times x}$, $y \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R}^y)$, and consider the PDE

$$F\left(\frac{\partial}{\partial v_1}, \frac{\partial}{\partial v_2}, \ldots, \frac{\partial}{\partial v_n}\right) x = y. \tag{S15}$$

in the variables $v_1, v_2, \ldots, v_n$. The problem addressed by the fundamental principle is to obtain necessary and sufficient conditions for the existence of a solution $x \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R}^x)$. This question can be viewed as a special case of finding conditions for the existence of a solution $x \in \mathbb{X}$ to the equation $F(x) = y$ with $F : \mathbb{X} \to \mathbb{Y}$ and $y \in \mathbb{Y}$. This question is, with all its algorithmic ramifications, one of the most central ones of mathematics. The fundamental principle gives a necessary and sufficient condition for solvability in the case of a linear constant-coefficient PDE.

Let $R \in \mathbb{R}[\xi]^{\bullet \times w}$, $M \in \mathbb{R}[\xi]^{\bullet \times l}$, and consider the ODE

$$R\left(\frac{d}{dt}\right) w = M\left(\frac{d}{dt}\right) \ell, \tag{S16}$$

containing both manifest variables w and latent variables ℓ. Define the *manifest behavior* as

$$\mathcal{B} = \{w \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^w) \mid \text{there exists } \ell \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^l) \text{ such that (S16) holds}\}.$$

The question addressed by the elimination theorem is: Does there exist a polynomial matrix $R' \in \mathbb{R}[\xi]^{\bullet \times w}$ such that $\mathcal{B}$ consists of exactly the $\mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^w)$-solutions of

$$R'\left(\frac{d}{dt}\right) w = 0?$$

In words, we wish to eliminate the latent variables ℓ from (S16). The full set of variables (w, ℓ) could be the variables w on the external terminals of a black box combined with the variables ℓ on the internal terminals of the subsystems (see Figure 1). The elimination theorem addresses the question as to whether there is an ODE containing only the external variables that describes the behavior of the variables on the external terminals. Although elimination is not possible in general for nonlinear differential systems, for linear time-invariant differential systems, elimination holds in full generality. The elimination theorem for linear constant-coefficient ODEs and PDEs turns out to be a straightforward consequence of the fundamental principle.

Let $N : \mathbb{Y} \to \mathbb{N}$ with $0 \in \mathbb{N}$. Obviously, a necessary condition for solvability of $F(x) = y$ is that $N \circ F = 0$ implies $N(y) = 0$. The problem is to turn this necessary condition into a sufficient one by first constructing a sufficiently rich family of maps N that annihilate F and verifying that these N's also annihilate y. Of course, the smaller, more concrete, and easier it is to compute the required set of maps N, the more useful the result.

Finding conditions for the solvability of (S15) is immediate if all maps N that annihilate F are considered. Clearly, $F(x) = y$ is solvable if and only if, for all maps $N : \mathbb{Y} \to \{0, 1\}$ such that $\mathtt{image}(N \circ F) = \{0\}$, it follows that $N(y) = 0$, since this statement is merely a cryptic way of stating that $y \in \mathtt{image}(F)$. The issue is to exploit additional structure on F so that it suffices to verify $N(y) = 0$ for a smaller, perhaps finite, set of Ns.

Let us illustrate what we are after in the case that F is a matrix. For $F \in \mathbb{R}^{y \times x}$ and $y \in \mathbb{R}^y$, a necessary and sufficient condition for the solvability of $Fx = y$ for $x \in \mathbb{R}^x$ is that y be in the span of the columns of F. Equivalently, every vector $n \in \mathbb{R}^y$ that annihilates F, meaning $n^\top F = 0$, ought to annihilate y, meaning $n^\top y = 0$. Equivalently, every element of a basis of the finite-dimensional linear space of n's such that $n^\top F = 0$ ought to be orthogonal to y. In this case it is thus trivial to see that there is a finite set of maps N that yield a necessary and sufficient condition for solvability.

Let us now turn to linear constant-coefficient PDEs. A necessary condition for solvability of (S15) is that all polynomial vectors $n \in \mathbb{R}[\xi_1, \xi_2, \ldots, \xi_n]^y$ that annihilate F, meaning $n^\top F = 0$, ought to annihilate y, meaning

$$n\left(\frac{\partial}{\partial v_1}, \frac{\partial}{\partial v_2}, \ldots, \frac{\partial}{\partial v_n}\right)^\top y = 0.$$

The fundamental principle states that this condition is also sufficient.

This result is effective because these conditions can be reduced to a finite set. Indeed, the set of polynomial vectors $n \in \mathbb{R}[\xi_1, \xi_2, \ldots, \xi_n]^y$ such that $n^\top F = 0$ is the left syzygy (see "Polynomial Modules and Syzygies'" of F, and, since it is a submodule of $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_n]^y$, it is finitely generated. Hence (37) is solvable if and only if

$$N\left(\frac{\partial}{\partial v_1}, \frac{\partial}{\partial v_2}, \ldots, \frac{\partial}{\partial v_n}\right) y = 0,$$

with the rows of N the transposes of a set of generators $n_1, \ldots n_p \in \mathbb{R}[\xi_1, \xi_2, \ldots, \xi_n]^y$ of the left syzygy of F. A set of generators of this syzygy can be computed using computer algebra. Note that the fundamental principle states that the image of the differential operator $F((\partial/\partial v_1), (\partial/\partial v_2), \ldots, (\partial/\partial v_n))$, acting on $\mathcal{C}^\infty$-functions, equals the kernel of $N((\partial/\partial v_1), (\partial/\partial v_2), \ldots, (\partial/\partial v_n))$. In particular, this

consequence of the fact that every LTIDS has a minimal kernel representation and that minimality is equivalent to the polynomial matrix in the kernel representation having full row rank. Hence, there never need be more differential equations than variables in a kernel representation of an LTIDS.

First principles models of interconnected systems often lead to a large tableau of equations, combining many algebraic equations with first-, second-, and higher order differential equations and involving many auxiliary variables, for example, resulting from the interconnection constraints.

result implies that every image of a linear constant-coefficient partial differential operator is a kernel of such an operator. However, it turns out that not every kernel is an image and that the kernels that are images correspond exactly to the controllable behaviors (see "Controllability and Image Representations").

The fundamental principle implies in particular that the scalar PDE

$$\pi\left(\frac{\partial}{\partial v_1}, \frac{\partial}{\partial v_2}, \ldots, \frac{\partial}{\partial v_{\mathtt{n}}}\right) x = f.$$

with $0 \neq \pi \in \mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$ and $f \in \mathcal{C}^\infty(\mathbb{R}^{\mathtt{n}}, \mathbb{R})$, always has a solution $x \in \mathcal{C}^\infty(\mathbb{R}^{\mathtt{n}}, \mathbb{R})$, a nontrivial result. For ODEs, this result is easy, since we know from linear system theory that the set of solutions of $\pi((d/dt))x = f$, where $\pi \in \mathbb{R}[\xi]$ is nonzero, is an affine subspace of dimension $\mathtt{degree}(\pi)$. There is indeed one solution for each initial condition

$$x(0), \frac{d}{dt}x(0), \ldots, \frac{d^{\mathtt{degree}(\pi)-1}}{dt^{\mathtt{degree}(\pi)-1}}x(0).$$

For PDEs, the existence of even one solution is far from evident.

The fundamental principle is immediate on a set-theoretic level, trivial for matrices, easy for ODEs, and deep for PDEs. The proof of the fundamental principle in the ODE case is an easy consequence of the Smith form. For PDEs the fundamental principle is an outcome of [S21]–[S23], where it is proven that $\mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$, viewed as a $\mathbb{R}((\partial/\partial_{v_1}), (\partial/\partial_{v_2}), \ldots, (\partial/\partial_{v_{\mathtt{n}}}))$ module, is "injective." The fundamental principle is valid, not only in the $\mathcal{C}^\infty$-case considered here but also for general distributions and for tempered distributions. However, the fundamental principle does not hold for compact support solutions or for $\mathcal{L}_2$ solutions. The solution space is relevant. Solvability of (S15) is not only a matter of algebra; it also involves analysis.

The elimination theorem follows from the fundamental principle in a straightforward way. The problem is to find all $w \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{w}})$ for which there exists $\ell \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{l}})$ such that $R(d/dt)w = M(d/dt)\ell$. By the fundamental principle, $R(d/dt)w$ belongs to the image of $M(d/dt)\ell$ if and only if

$$N\left(\frac{d}{dt}\right)^\top R\left(\frac{d}{dt}\right) w = 0,$$

where the transposes of the rows of N form a set of generators of the left syzygy of M. The elimination theorem follows, since we have pinpointed the required polynomial matrix $R' = NR$. Algorithmically, therefore, elimination amounts to finding a set of generators of the left syzygy of M, a standard element in computer algebra packages.

The elimination theorem and its proof are valid for linear constant-coefficient PDEs. The elimination theorem for PDEs implies, for example, that the electrical variables $\vec{E}, \vec{j}, \rho$ in Maxwell's equations obey, after elimination of the magnetic field $\vec{B}$, a system of linear constant-coefficient PDEs (see the section "PDEs" for the PDE that is obtained after elimination of $\vec{B}$).

Elimination questions play an important role in mathematics, for example, to determine properties of the projection of a subset $\mathcal{A} \subseteq \mathbb{R}^{\mathtt{n}} \times \mathbb{R}^{\mathtt{m}}$ onto the first $\mathtt{n}$ coordinates. Openness, linearity, and convexity of $\mathcal{A}$ are properties that are preserved under projection, while closedness, being an algebraic variety, and being a differentiable manifold are not necessarily preserved under projection. The Tarski-Seidenberg theorem states that the projection of a semi-algebraic set is also a semi-algebraic set. A subset of $\mathbb{R}^{\mathtt{n}}$ is *semi-algebraic* if it is a finite union of sets of the form

$$\{x \in \mathbb{R}^{\mathtt{n}} \mid p_0(x) = 0 \text{ and } p_k(x) > 0 \text{ for } k = 1, 2, \ldots, d\},$$

with $p_0, p_1, \ldots, p_d \in \mathrm{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$.

The elimination theorem can be phrased in terms of projections. Elimination means projecting a subset, in the case of linear constant-coefficient ODEs or PDEs, onto a subset of $\mathcal{C}^\infty(\mathbb{R}^{\mathtt{n}}, \mathbb{R}^{\mathtt{m}+\mathtt{p}})$ on $\mathcal{C}^\infty(\mathbb{R}^{\mathtt{n}}, \mathbb{R}^{\mathtt{m}+\mathtt{p}})$, and the question is whether properties of the projected set are inherited by the projection. The elimination theorem states that, if the projected set consists of the $\mathcal{C}^\infty$-solutions of a system of linear constant-coefficient ODEs or PDEs, then the projection also consists of the $\mathcal{C}^\infty$-solutions of a system of linear constant-coefficient ODEs or PDEs.

The fundamental principle and the elimination theorem play a basic role in *n*D-system theory [43], [S16], [S17]. The emphasis is usually on $\mathcal{C}^\infty$ or the distributions as the basic solution space for the PDEs. In [S25], generalization to alternative solution spaces is discussed.

REFERENCES

[S22] L. Ehrenpreis, *Fourier Analysis in Several Complex Variables*, New York: Wiley-Interscience, 1970 .

[S23] B. Malgrange, "Systèmes différentiels à coefficients constants," *Séminaire Bourbaki*, vol. 8, pp. 79–89, 1964.

[S24] V.P. Palamodov, *Linear Differential Operators with Constant Coefficients*, New York: Springer Verlag, 1970.

[S25] S. Shankar, "Geometric completeness of distribution spaces," *Acta Applicandae Mathematicae*, vol. 77, pp. 163–180, 2003.

The elimination theorem guarantees that, for LTIDSs, this tableau of DAEs can be collapsed to a number of differential equations that is at most equal to the number of manifest variables. For example, there are in total 26 equations that describe the RLC circuit of Figure 8, namely, 12 module equations, 12 interconnection equations, and two equations for the manifest variable assignment. These equations contain 28 latent variables, namely, the terminal voltages and currents, as well as two manifest variables, namely, the port voltage and current. The manifest behavior obtained after eliminating the latent variables (see [4, pp. 11–12]) consists of one scalar differential equation.

It follows from the elimination theorem that $\mathcal{L}^\bullet$ has nice properties. $\mathcal{L}^\bullet$ is closed under addition, intersection, and projection, as well as under action and inverse action of a linear constant-coefficient differential operator. Although closure under addition is not as straightforward as it may seem, it is a simple consequence of the elimination theorem. Closure under addition can be proven as follows. Let $\mathcal{B}_1$ be the behavior of $R_1(d/dt)w = 0$, and let $\mathcal{B}_2$ be the behavior of $R_2(d/dt)w = 0$. Then $\mathcal{B}_1 + \mathcal{B}_2$ is the w behavior of $R_1(d/dt)w_1 = 0$, $R_2(d/dt)w_2 = 0, w = w_1 + w_2$. Now eliminate w_1 and w_2, and conclude that $\mathcal{B}_1 + \mathcal{B}_2 \in \mathcal{L}^\bullet$.

Controllability and Image Representations

The notions of controllability and stabilizability are discussed in the section "Controllability as a System Property." For LTIDSs, the following results provide concrete tests for verifying controllability and stabilizability.

Proposition 3 (Controllability Test)

The system (25) defines a controllable system if and only if the rank of $R(\lambda)$ is the same for all $\lambda \in \mathbb{C}$.

Proposition 4. (Stabilizability Test)

The system (25) defines a stabilizable system if and only if the rank of $R(\lambda)$ is the same for all complex λ such that $\texttt{real}\,(\lambda) \geq 0$.

LTIDSs admit, by definition, a kernel representation (25). First principles modeling usually leads to equations with latent variables, such as (27). For LTIDSs, the elimination theorem implies that latent variables can be eliminated, leading to a kernel representation for the manifest behavior. An interesting family of differential systems appears to be lacking in this classification, namely, those described by

$$w = M\left(\frac{d}{dt}\right)\ell, \tag{29}$$

with w the manifest variables, and ℓ the latent variables. Note that this system of differential equations leaves the latent variables ℓ free. Since the manifest behavior $\mathcal{B}$ of (29) equals $\texttt{image}\,(M(d/dt))$, such a representation is an *image representation* of the manifest behavior. Of course, since (29) is a special case of a system with latent variables, it follows from the elimination theorem that its manifest behavior $\mathcal{B}$ can be described by a kernel representation (25) for a suitable R. Therefore, for constant-coefficient linear differential operators, every image is a kernel. The dual question concerns whether every kernel is also an image, and, if not, what additional properties need to hold for representability of a kernel as an image. This question is a compelling one from a mathematical point of view. Surprisingly, this question also characterizes an essential system-theoretic property, namely, controllability!

Theorem 5 (Image Representation)

$\mathcal{B} \in \mathcal{L}^\bullet$ admits an image representation (29) if and only if it is controllable.

The controllability test given by Proposition 3 does not appear to be practical, since it involves, at least in principle, checking the rank of $R(\lambda)$ for all $\lambda \in \mathbb{C}$, an infinite set of matrices. However, representability as an image leads to a more practical test for controllability. We wish to determine whether (25) defines a controllable system. The algorithm involves the following steps, each of which is a standard problem in computer algebra (see "Polynomial Modules and Syzygies" for the nomenclature used):

1) Compute a set of generators of the right syzygy of the rows of R, that is, compute a matrix $M \in \mathbb{R}[\xi]^{\mathtt{W}\times\bullet}$ such that the columns of M span the module of polynomial vectors $a \in \mathbb{R}[\xi]^{\mathtt{W}}$ such that $Ra = 0$.
2) Compute the left syzygy of M, that is, compute a matrix $R' \in \mathbb{R}[\xi]^{\bullet\times\mathtt{W}}$ such that the rows of R' span the module of polynomial vectors $b \in \mathbb{R}[\xi]^{1\times\mathtt{W}}$ satisfying $bM = 0$.
3) Check whether the modules generated by the rows of R and the rows of R' are equal. The system defined by (25) is controllable if and only if the modules generated by the rows of R and R' are equal.

The equivalence of existence of an image representation and controllability expressed by Theorem 5 brings the importance of image representations in linear system theory to the foreground [1, p. 567]. Image representations are also used under the name *differentially flat* systems [12]. Right coprime factorizations of transfer functions can be interpreted as image representations. Theorem 5 generalizes to linear constant-coefficient PDEs [34] (see the section "PDEs") as well as to linear delay-differential systems [35], [36], but the proof is much harder in these cases.

The Controllable Part of a LTIDS

Controllability plays a central role in the theory of LTIDSs, just as state controllability does for input/state/output systems. In the remainder of this section, some further results involving the controllable part of a system and the sum decomposition of a behavior into a controllable and an autonomous part are given.

Let $\mathcal{B} \in \mathcal{L}^\bullet$. The *controllable part* of $\mathcal{B}$, denoted by $\mathcal{B}_{\text{controllable}}$ can be defined in many equivalent ways. The simplest approach is to define $\mathcal{B}_{\text{controllable}}$ to mean the

largest controllable subsystem of $\mathcal{B}$. $\mathcal{B}_{\text{controllable}}$ is hence defined by i) $\mathcal{B}_{\text{controllable}} \in \mathcal{L}^{\bullet}$, ii) $\mathcal{B}_{\text{controllable}} \subseteq \mathcal{B}$, iii) $\mathcal{B}_{\text{controllable}}$ is controllable, and iv)

$$\mathcal{B}' \in \mathcal{L}^{\bullet}, \mathcal{B}' \text{ controllable, and } \mathcal{B}' \subseteq \mathcal{B}$$

imply

$$\mathcal{B}' \subseteq \mathcal{B}_{\text{controllable}}.$$

It is easily proven that $\mathcal{B}_{\text{controllable}}$ exists. In fact, it is possible to compute a kernel representation of $\mathcal{B}_{\text{controllable}}$ from a kernel representation of $\mathcal{B}$ by carrying out the first step of the algorithm for checking controllability. The polynomial matrix M computed in the first step of this algorithm yields the image representation $\mathcal{B}_{\text{controllable}} = \texttt{image}(M(d/dt))$. The polynomial matrix R' computed in the second step yields the kernel representation $\mathcal{B}_{\text{controllable}} = \texttt{kernel}(R'(d/dt))$.

Every system $\mathcal{B} \in \mathcal{L}^{\bullet}$ admits a minimal kernel representation (25), that is, a kernel representation (25) for which R has full row rank. Recall that $P \in \mathbb{R}[\xi]^{\bullet\times\bullet}$ is *left prime* over $\mathbb{R}[\xi]$ if $P = FP'$, with $F, P' \in \mathbb{R}[\xi]^{\bullet\times\bullet}$ and F square, implies that F is unimodular. The full row rank polynomial matrix R can be factored as $R = FR'$ with F square and R' left prime over $\mathbb{R}[\xi]$. The system $R'(d/dt)w = 0$ defines the controllable part of $\mathcal{B}$. A minimal kernel representation (25) defines a controllable system if and only if R is left prime over $\mathbb{R}[\xi]$.

At the other extreme from controllability are autonomous systems. $\mathcal{B} \in \mathcal{L}^{\bullet}$ is *autonomous* if

$$w_1, w_2 \in \mathcal{B} \quad \text{and} \quad w_1(t) = w_2(t) \quad \text{for all} \quad t < 0$$

implies

$$w_1 = w_2.$$

$\mathcal{B} \in \mathcal{L}^{\bullet}$ is autonomous if and only if it admits a kernel representation (25) with R square and $\det(R) \neq 0$.

The *characteristic polynomial* $\chi_{\mathcal{B}}$ of an element $\mathcal{B} \in \mathcal{L}^{\bullet}$ is defined as follows. If $\mathcal{B}$ is autonomous, then $\mathcal{B}$ has a kernel representation (25) with R square, $\det(R) \neq 0$, and $\det(R)$ monic. Define $\chi_{\mathcal{B}} := \det(R)$. If $\mathcal{B}$ is not autonomous, define $\chi_{\mathcal{B}} := 0$.

A system $\mathcal{B} \in \mathcal{L}^{\mathbb{W}}$ is *stable* if

$$w \in \mathcal{B}$$

implies

$$w(t) \to 0 \text{ as } t \to \infty.$$

Obviously, for $\mathcal{B} \in \mathcal{L}^{\mathbb{W}}$ to be stable, it must be autonomous, since nonautonomous systems $\mathcal{B} \in \mathcal{L}^{\mathbb{W}}$ contain free variables, which prevent the conclusion that $w(t) \to 0$ as $t \to \infty$ for all $w \in \mathcal{B}$. In fact, $\mathcal{B}$ is stable if and only if its characteristic polynomial $\chi_{\mathcal{B}}$ is Hurwitz, that is, if and only if all of its roots have negative real part. We consider stability as a property of autonomous systems. There are many other forms of stability that are of interest, related to dissipativity, or obtained by setting certain variables of $\mathcal{B}$ to zero, but here we use the above definition of stability, sometimes called *Lyapunov stability*.

Let $\mathcal{B} \in \mathcal{L}^{\bullet}$. Then $\mathcal{B}$ admits a direct sum decomposition $\mathcal{B} = \mathcal{B}_1 \oplus \mathcal{B}_2$, with $\mathcal{B}_1 \in \mathcal{L}^{\bullet}$ controllable and $\mathcal{B}_2 \in \mathcal{L}^{\bullet}$ autonomous. This sum decomposition is not unique but enjoys many invariant properties. For example, $\mathcal{B}_1$ is unique and is, in fact, equal to the controllable part $\mathcal{B}_{\text{controllable}}$ of $\mathcal{B}$, while the characteristic polynomial of the complementary autonomous system $\mathcal{B}_2$ is the same for all $\mathcal{B}_2$ such that $\mathcal{B} = \mathcal{B}_1 \oplus \mathcal{B}_2$. The system $\mathcal{B} \in \mathcal{L}^{\bullet}$ is stabilizable if and only if the characteristic polynomial of the complementary autonomous part of $\mathcal{B}$ is Hurwitz.

A system is controllable if and only if it is equal to its controllable part. Two systems have the same controllable part if and only if they have the same compact support behavior. More precisely, consider $\mathcal{B} \in \mathcal{L}^{\bullet}$, and define $\mathcal{B}_{\text{compact}} := \{w \in \mathcal{B} \mid w \text{ has compact support}\}$. Then

$$\mathcal{B}_{1,\text{compact}} = \mathcal{B}_{2,\text{compact}}$$

if and only if

$$\mathcal{B}_1 \text{ and } \mathcal{B}_2 \text{ have the same controllable part.}$$

In particular, therefore, the compact support behavior determines only the controllable part of a system. A similar result also holds for the $\mathcal{L}_2$-behavior. Specifically,

$$\mathcal{B}_1 \cap \mathcal{L}_2(\mathbb{R}, \mathbb{R}^{\bullet}) = \mathcal{B}_2 \cap \mathcal{L}_2(\mathbb{R}, \mathbb{R}^{\bullet})$$

if and only if

$$\mathcal{B}_1 \text{ and } \mathcal{B}_2 \text{ have the same controllable part.}$$

RATIONAL REPRESENTATIONS OF LTIDSs

LTIDSs admit many representations. Depending on the context, some representations are more useful than others. Kernel representations (25) provide the basic definition of an LTIDS and, as such, these representations play the central role in theoretical developments. Latent variable representations (27) are usually the end result of a first principles modeling process, and image representations (29) characterize controllability. Alternative useful system representations are input/output, state, and input/state/output representations. These representations play a central role in linear system theory.

In this section we discuss representations with rational symbols. These representations place the notion of a transfer function firmly into the setting of behavioral systems. Although transfer functions are usually defined in terms of Laplace transforms, we avoid Laplace transforms altogether.

Thinking of a dynamical system as a behavior, and of interconnection as variable sharing, gets the physics right.

A transfer function is looked on in the spirit of Heaviside's symbolic calculus, in which a polynomial or a rational function is viewed in terms of indeterminates, for which a differential operator, shift, or delay operator can be substituted. Consequently, the development is unencumbered by domains of convergence and exponential growth of signals. However, in all fairness, we point out that we limit the discussion to transfer functions that are rational. For rational transfer functions it is much easier to bypass Laplace transforms than for irrational ones.

Rational Symbols

For $R \in \mathbb{R}(\xi)^{\bullet\times\mathbf{W}}$, a matrix of rational functions, we wish to determine what is meant by the set of solutions of the system

$$R\left(\frac{d}{dt}\right)w = 0. \tag{30}$$

Since R is not polynomial, it is not clear when $w : \mathbb{R} \to \mathbb{R}^{\mathbf{W}}$ is a solution of (30). This ambiguity is not a question of smoothness, but rather is a matter of assigning a meaning to the equality (30), since $R(d/dt)$ is not a differential operator.

Let $M \in \mathbb{R}(\xi)^{\bullet\times\bullet}$. A *left polynomial matrix factorization* of M is a pair (P, Q) such that $P, Q \in \mathbb{R}[\xi]^{\bullet\times\bullet}$, P is square with $\mathtt{det}(P) \neq 0$, and $M = P^{-1}Q$. The left polynomial matrix factorization (P, Q) of M is *left coprime* if, in addition, the polynomial matrix $[P \;\; Q]$ is left prime over $\mathbb{R}[\xi]$. The pair of polynomial matrices (P, Q) is *left prime* over $\mathbb{R}[\xi]$ if $[P \;\; Q] = F[P' \;\; Q']$, with $F, P', Q' \in \mathbb{R}[\xi]^{\bullet\times\bullet}$ implies that F is $\mathbb{R}[\xi]$-unimodular.

Equation (30) can be reduced to an ODE as follows. Let (P, Q) be a left coprime polynomial matrix factorization of $R = P^{-1}Q$. The map $w : \mathbb{R} \to \mathbb{R}^{\mathbf{W}}$ is defined to be a solution of (30) if $Q(d/dt)w = 0$. Denote the set of solutions of (30) by $\mathtt{kernel}\,(R(d/dt))$. Hence (30) defines the LTIDS $\Sigma = (\mathbb{R}, \mathbb{R}^{\mathbf{W}}, \mathtt{kernel}(R(d/dt))) := (\mathbb{R}, \mathbb{R}^{\mathbf{W}}, \mathtt{kernel}(Q(d/dt))) \in \mathcal{L}^{\mathbf{W}}$. R is the *symbol* that defines this system as a kernel representation. Note that it follows from this definition of a solution of (30) that, for $F \in \mathbb{R}(\xi)^{\bullet\times\bullet}$, $F(d/dt)$ is a point-to-set map since there can be many solutions $w_2 \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet)$ to $w_2 = F(d/dt)w_1$ corresponding to a given $w_1 \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet)$. It is in this point-to-set interpretation of $R(d/dt)$ that $\mathtt{kernel}(R(d/dt))$ is interpreted. $\mathtt{kernel}(R(d/dt))$ consists of the $w \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathbf{W}})$ such that $0 \in R(d/dt)w$.

For example, for $R = q/p$, with $p, q \in \mathbb{R}[\xi]$ coprime, meaning no common roots, the set of solutions of $q/p(d/dt)w = 0$ is defined to be equal to that of the differential equation $q(d/dt)w = 0$. In this example, the behavior is finite dimensional, with dimension equal to $\mathtt{degree}(q)$. Usually, R is wide, that is, R has more columns that rows. For example, the behavior of $(q_1/p_1)(d/dt)w_1 + (q_2/p_2)(d/dt)w_2 = 0$, with $p_1, q_1 \in \mathbb{R}[\xi]$ and $p_2, q_2 \in \mathbb{R}[\xi]$ both coprime, is equal to the set of solutions of $p_2'q_1(d/dt)w_1 + p_1'q_2(d/dt)w_2 = 0$, where $p_1 = dp_1'$, $p_2 = dp_2'$, $d \in \mathbb{R}[\xi]$, and $p_1', p_2' \in \mathbb{R}[\xi]$ are coprime. Hence, because of common factors, the behavior of $G_1(d/dt)w_1 = G_2(d/dt)w_2$ with $G_1, G_2 \in \mathbb{R}(\xi)$, $G_1 \neq 0$, is not necessarily equal to the behavior of $w_1 = (G_1^{-1}G_2)(d/dt)w_2$.

The definition of the behavior defined by (30) can be justified in many ways, for example, from the following state-space point of view. Decompose R into $R = S + P$ with $S \in \mathbb{R}(\xi)^{\bullet\times\mathbf{W}}$ strictly proper, and $P \in \mathbb{R}[\xi]^{\bullet\times\mathbf{W}}$ polynomial. Let $S(\xi) = C(I\xi - A)^{-1}B$, with (A, B) controllable. This decomposition leads to the system

$$\frac{d}{dt}x = Ax + Bw, \quad 0 = Cx + P\left(\frac{d}{dt}\right)w.$$

The manifest behavior of this system, with x viewed as a latent variable, is precisely the set of solutions of (30), as defined by $\mathtt{kernel}(R(d/dt))$.

The notion of annihilator of a behavior is readily generalized from the differential case discussed in the section "Differential Annihilators" to annihilators that involve vectors of rational functions. These annihilators are studied in [37]. Noteworthy is the result [37, thm. 11] that the set of rational annihilators of $\mathcal{B} \in \mathcal{L}^{\mathbf{W}}$ forms an $\mathbb{R}(\xi)$-vector subspace of $\mathbb{R}(\xi)^{\mathbf{W}}$ if and only if $\mathcal{B}$ is controllable.

Since the representations (25) are a subset of the representations (30), matrices of rational functions yield yet another class of representations of $\mathcal{L}^\bullet$. This set of representations is richer and more redundant than the polynomial matrices and offers new possibilities. For example, representations with rational symbols lead to norm-preserving kernel and image representations.

Transfer Functions

Obviously,

$$y = G\left(\frac{d}{dt}\right)u, w = (u, y), \tag{31}$$

with $G \in \mathbb{R}(\xi)^{\bullet\times\bullet}$ is a special case of (30) with $R = [G \;\; -I]$ and $w = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right]$. By definition, therefore, (31) defines the LTIDS whose behavior is the set of solutions of

$$P\left(\frac{d}{dt}\right)y = Q\left(\frac{d}{dt}\right)u,$$

with (P, Q) a left coprime polynomial matrix factorization of $G = P^{-1}Q$. It follows that (31) defines a controllable system. Consequently, transfer functions always define controllable systems. This fact is good news and bad news. The good news is that LTIDSs defined by transfer functions automatically have desirable properties, for example, they are always controllable and hence stabilizable. The bad news is that transfer functions are incapable of modeling uncontrollable behaviors, for example, autonomous systems, which are often used as physical models. Many control problems, such as disturbance decoupling, aim at making systems uncontrollable.

Transfer functions are also unable to deal with bad properties that emerge by interconnection. For example, a series connection of systems can lead to a lack of controllability—this fact is also obvious from state-space thinking—and a series connection of scalar transfer functions does not commute. More precisely, for $G_1, G_2 \in \mathbb{R}(\xi)$, the (u, y)-behaviors defined by

$$v = G_1\left(\frac{d}{dt}\right)u, \quad y = G_2\left(\frac{d}{dt}\right)v,$$
$$z = G_2\left(\frac{d}{dt}\right)u, \quad y = G_1\left(\frac{d}{dt}\right)z,$$

and

$$y = G_2\left(\frac{d}{dt}\right)G_1\left(\frac{d}{dt}\right)u$$

may all be different. For example, $G_1(\xi) = (\xi/\xi + 1)$ followed by $G_2(\xi) = (\xi + 1/\xi)$ yields the behavioral equations $(d/dt + 1)y_1 = (d/dt)u$, $(d/dt)y = (d/dt + 1)u_2$, $y_1 = u_2$. After elimination of u_1, y_1, u_2, and y_2, we obtain the behavior $(d/dt)y = (d/dt)u$. However, $G_2(\xi)$ followed by G_1 yields $(d/dt)y_1 = ((d/dt) + 1)u$, $(d/dt)u_2 = ((d/dt) + 1)y$, $y_1 = u_2$. After elimination of u_1, y_1, u_2, and y_2, we obtain the behavior $((d/dt) + 1)y = u((d/dt) + 1)$. The behavior defined $y = G_2(d/dt)G_1(d/dt)u$ is obviously $y = u$, since the rational function $(\xi/\xi + 1)(\xi + 1/\xi) = 1$. Both series connections are not controllable, although they have the same controllable part, namely, $y = u$, the behavior of $y = G_2(d/dt)G_1(d/dt)u$. The first series connection also admits the responses $y(t) = u(t) + c$ with $c \in \mathbb{R}$ an arbitrary constant, and the second series connection also admits the responses $y(t) = u(t) + ce^{-t}$ with $c \in \mathbb{R}$ an arbitrary constant.

Consider the system

$$P\left(\frac{d}{dt}\right)y = Q\left(\frac{d}{dt}\right)u, \quad w = (u, y), \tag{32}$$

where $P, Q \in \mathbb{R}[\xi]^{\bullet\times\bullet}$, P is square, and $\det(P) \neq 0$. Let $\mathcal{B}$ be its behavior. In this behavior, u and y have special properties. In particular, u is free, meaning that for all $u \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet)$, there exists $y \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet)$ such that $(u, y) \in \mathcal{B}$. Moreover, given $u \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet)$, the set of $y \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^\bullet)$ such that $(u, y) \in \mathcal{B}$ is a finite dimensional linear variety. This structure implies that the future of y is uniquely specified by $u, \mathcal{B}$, and the past of y. These properties are often taken as the defining properties for input and output. See "Cause and Effect" for some remarks on the properties of (32) with respect to nonanticipation. The matrix $P^{-1}Q$ of rational functions is called the *transfer function* of (32).

Now consider the systems

$$P_1\left(\frac{d}{dt}\right)y = Q_1\left(\frac{d}{dt}\right)u \tag{33}$$

and

$$P_2\left(\frac{d}{dt}\right)y = Q_2\left(\frac{d}{dt}\right)u, \tag{34}$$

where $P_1, P_2, Q_1, Q_2 \in \mathbb{R}[\xi]^{\bullet\times\bullet}$, P_1, P_2 are square, $\det(P_1) \neq 0$, and $\det(P_2) \neq 0$. Assume that these systems have the same transfer function, that is, $P_1^{-1}Q_1 = P_2^{-1}Q_2$. What does this equality mean as far as the associated behaviors are concerned?

Theorem 6 (Controllability and Transfer Functions)

The systems (33) and (34) have the same transfer function if and only if they have the same controllable part. Moreover, given $G \in \mathbb{R}(\xi)^{\bullet\times\bullet}$, there is exactly one controllable system (32) that has transfer function G.

The manifest behavior of

$$w = M\left(\frac{d}{dt}\right)\ell, \tag{35}$$

with $M \in \mathbb{R}(\xi)^{\bullet\times\bullet}$, is always controllable, as is the case for polynomial representations. Hence, image$(M(d/dt))$ defines a controllable LTIDS. This image must be interpreted properly since $M(d/dt)$ is a point-to-set map.

The power of rational representations becomes evident when we consider norm-preserving representations. Let $\mathcal{B} \in \mathcal{L}^\bullet$. Define the $\mathcal{L}_2$*-behavior* as $\mathcal{B}_{\mathcal{L}_2} := \{w \in \mathcal{B} \cap \mathcal{L}_2(\mathbb{R}, \mathbb{R}^\bullet)\}$. Two systems in $\mathcal{L}^\bullet$ have the same $\mathcal{L}_2$-behavior if and only if they have the same controllable part. Consider the system $\Sigma = (\mathbb{R}, \mathbb{R}^{\mathbf{n}_1} \times \mathbb{R}^{\mathbf{n}_2}, \mathcal{B})$. This system is $\mathcal{L}_2$*–norm preserving* if

$$(w_1, w_2) \in \mathcal{B}_{\mathcal{L}_2}$$

implies

$$\|w_1\|_{\mathcal{L}_2(\mathbb{R}, \mathbb{R}^{\mathbf{n}_1})} = \|w_2\|_{\mathcal{L}_2(\mathbb{R}, \mathbb{R}^{\mathbf{n}_2})}.$$

The system defined by

$$w_1 = F\left(\frac{d}{dt}\right)w_2, \tag{36}$$

with $F \in \mathbb{R}(\xi)^{\mathbf{n}_1\times\mathbf{n}_2}$, is $\mathcal{L}_2$-norm preserving if and only if

$$F(-i\omega)^\top F(i\omega) = I_{\mathbf{n}_1} \quad \text{for all} \quad \omega \in \mathbb{R}.$$

Note that if this system (36) is $\mathcal{L}_2$-norm preserving then F is rational but not polynomial. Rational representations lead to the possibility of obtaining $\mathcal{L}_2$-norm preserving Rs in (30) and Ms in (35). It can, in fact, be shown that $\mathcal{B} \in \mathcal{L}^\bullet$ allows an $\mathcal{L}_2$-norm preserving representation (30) if and only if the characteristic polynomial of the autonomous part of $\mathcal{B}$ has no roots on the imaginary axis. $\mathcal{B} \in \mathcal{L}^\bullet$ allows a representation (30) with R a matrix of proper rational functions without poles in the closed right half of the complex plane and $\mathcal{L}_2$-norm preserving if and only if $\mathcal{B}$ is stabilizable. $\mathcal{B}$ allows a representation (35) with M a matrix of proper rational functions without poles in the closed right half of the complex plane and $\mathcal{L}_2$-norm preserving if and only if $\mathcal{B}$ is controllable. Such representations are useful for model reduction of unstable systems [38].

IMPLEMENTABILITY OF CONTROLLED BEHAVIORS FOR LTIDSs

In the present section, the control configuration discussed in the section "Control as Interconnection" and illustrated in Figure 15 is considered for the case in which the plant and controller are both LTIDSs. Denote by $\mathtt{v}$ the number of to-be-controlled variables and by $\mathtt{k}$ the number of control variables. The control variables are the variables through which the plant interacts with the controller. In the case of a feedback controller these variables include the sensor outputs as well as the control inputs. Control variables should not be confused with control inputs.

The plant has the behavior $\mathcal{B}_{\text{plant}} \in \mathcal{L}^{\mathtt{v}+\mathtt{k}}$. This behavior involves the to-be-controlled variables as well as the control variables. The controller restricts the control variables and has the behavior $\mathcal{K} \in \mathcal{L}^{\mathtt{k}}$. The controlled behavior, obtained by letting the controller act on the plant, is

$$\mathcal{B}_{\text{controlled}} = \{v \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{v}}) \mid \text{there exists } k \in \mathcal{K} \text{ such that } (v, k) \in \mathcal{B}_{\text{plant}}\}.$$

By the elimination theorem, $\mathcal{B}_{\text{controlled}} \in \mathcal{L}^{\mathtt{v}}$. The control problem is to choose, for a given plant $\mathcal{B}_{\text{plant}}$, a controller $\mathcal{K}$ from a set of admissible controllers, such that $\mathcal{B}_{\text{controlled}}$ meets certain specifications, or is optimal in some sense.

In this section, we assume that all LTIDS controllers $\mathcal{K} \in \mathcal{L}^{\mathtt{k}}$ are admissible. The following question emerges: *Given a plant* $\mathcal{B}_{\text{plant}} \in \mathcal{L}^{\mathtt{v}+\mathtt{k}}$, *which controlled behaviors* $\mathcal{B}_{\text{controlled}} \in \mathcal{L}^{\mathtt{v}}$ *can be obtained by choosing* $\mathcal{K} \in \mathcal{L}^{\mathtt{k}}$*?* The controlled behavior $\mathcal{B}_{\text{controlled}}$ is *implementable* if such a $\mathcal{K}$ exists, and $\mathcal{K}$ is then said to *implement* $\mathcal{B}_{\text{controlled}}$.

Implementable behaviors can be completely characterized as all behaviors that are wedged in between two extreme behaviors, the *hidden* plant behavior $\mathcal{N}$ (hidden because $\mathcal{N}$ consists of the v-trajectories that are compatible with the control signal $k = 0$), and the *unrestricted* plant behavior $\mathcal{P}$ (unrestricted because $\mathcal{P}$ consists of the v-trajectories that can occur without restrictions imposed on k by the controller). In the section "Control as Interconnection," $\mathcal{P}$ is denoted as $\mathcal{B}_{\text{uncontrolled}}$. $\mathcal{N}$ and $\mathcal{P}$ are defined by

$$\mathcal{N} := \{v \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{v}}) \mid (v, 0) \in \mathcal{B}_{\text{plant}}\}$$

and

$$\mathcal{P} := \{v \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{v}}) \mid \text{there exists } k \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{v}}) \text{ such that } (v, k) \in \mathcal{B}_{\text{plant}}\}.$$

Note that $\mathcal{N}$ and $\mathcal{K}$ are implementable behaviors, implemented, respectively, by the most stringent controller $\mathcal{K} = \{0\} \in \mathcal{L}^{\mathtt{k}}$ and the most liberal controller $\mathcal{K} = \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{\mathtt{k}}) \in \mathcal{L}^{\mathtt{k}}$.

The following theorem characterizes the implementable behaviors.

Theorem 7 (Implementability Theorem)

$\mathcal{B}_{\text{controlled}} \in \mathcal{L}^{\mathtt{v}}$ is implementable by some $\mathcal{K} \in \mathcal{L}^{\mathtt{k}}$ if and only if

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

That every implementable controlled behavior must be contained in $\mathcal{P}$ and contains $\mathcal{N}$ is obvious. The implementability theorem states that these conditions are the only constraints imposed on the controlled behavior, provided every controller $\mathcal{K} \in \mathcal{L}^{\mathtt{k}}$ is admissible. The implementability theorem shows that, for LTIDSs and without constraints on the admissible controllers, we can think of control design as finding a behavior that is wedged in between two given behaviors. Mathematically, this characterization of the implementable controlled behavior is simple and intuitive.

As a particular application of the implementability theorem, consider the case in which the to-be-controlled variables v are observable from the control variables k in the plant $\mathcal{B}_{\text{plant}}$. Observability of v from k means that

$$(v_1, k), (v_2, k) \in \mathcal{B}_{\text{plant}}$$

implies

$$v_1 = v_2.$$

Observability is equivalent to $\mathcal{N} = \{0\}$. It follows from the implementability theorem that, in the case in which v is observable from k, every behavior contained in $\mathcal{P}$ is implementable, and control reduces to choosing a sub-behavior of the uncontrolled behavior. Next, consider the case in which the to-be-controlled variables v are detectable from the control variables k in the plant $\mathcal{B}_{\text{plant}}$. Detectability of v from k means that

$$(v_1, k), (v_2, k) \in \mathcal{B}_{\text{plant}}$$

implies

$$v_1(t) - v_2(t) \to 0 \quad \text{as} \quad t \to \infty.$$

Detectability of v from k is equivalent to requiring $\mathcal{N}$ to be stable. It follows from the implementability theorem that there exists a controller $\mathcal{K} \in \mathcal{L}^{\mathbf{k}}$ such that $\mathcal{B}_{\text{controlled}}$ is stable if and only if the to-be-controlled variables v are detectable from the control variables k in the plant $\mathcal{B}_{\text{plant}}$.

Often, in applications, not all controllers $\mathcal{K} \in \mathcal{L}^{\mathbf{k}}$ are admissible. For example, $\mathcal{K}$ may be required to be dissipative or satisfy information structure constraints, or, if some of the control variables are sensor outputs, it is natural to require that $\mathcal{K}$ leave these variables free. Or, when $\mathcal{B}_{\text{plant}}$ contains exogenous variables, such as disturbances or command signals, $\mathcal{K}$ may be required to be such that the behavior of these signals remains unchanged in $\mathcal{B}_{\text{controlled}}$. For control as interconnection the issue of determining a reasonable class of admissible controllers is subtle [39]–[41].

PDEs

Models using PDEs dominate physics. In engineering, such models are of great importance for spatially distributed systems arising, for example, in electromagnetics, as models for waveguides and semiconductor devices, and in the study of materials and flexible structures. In control theory, systems described by PDEs have not attained a central role, and they are not part of the standard curriculum. PDE models are often formalized as *infinite-dimensional* systems, which include time-delay systems, leading to state models with an infinite-dimensional state space. A rich mathematical theory exists for such systems, based on functional analysis with semigroups and unbounded operators [42]. While this point of view can be appreciated because of the special role that time plays in many applications, it has several disadvantages as well. PDEs serve to describe phenomena involving space and time. Moreover, models involving PDEs often allow a simple algebraic parameterization using polynomials in several variables. The time-space interplay, as well as the algebraic structure, become opaque in the semigroup setting.

By considering both time and space as independent variables, these shortcomings can, to a large extent, be avoided. In this section, we briefly explain how behavioral theory can be adapted to deal with nD systems, in particular, with constant coefficient linear PDEs, as developed in [34] and [43]–[46]. For simplicity, only constant coefficient linear PDEs without boundary conditions are discussed, with Maxwell's equations in free space as an example.

The independent variables in Maxwell's equations are (t, x, y, z), time and space. The dependent variables are the electric field $\vec{E}$, the magnetic field $\vec{B}$, the current density $\vec{j}$, and the charge density ρ. The behavioral equations are

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

where $\nabla\cdot$ and $\nabla\times$ denote the divergence and curl, respectively, ε_0 is the dielectric constant of free space, and c is the speed of light in free space. We think of Maxwell's equations as a kernel representation for the fields $\vec{E} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$, $\vec{B} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$, $\vec{j} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$, and $\rho : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$, which, according to Maxwell's equations, can occur in free space.

A *linear shift-invariant* nD *differential system* is defined as $\Sigma = (\mathbb{R}^{\mathbf{n}}, \mathbb{R}^{\mathbf{w}}, \mathcal{B})$, with $\mathbb{R}^{\mathbf{n}}$ the set of *independent* variables, $\mathbb{R}^{\mathbf{w}}$ the set of *dependent* variables, and with the behavior $\mathcal{B} \subseteq (\mathbb{R}^{\mathbf{w}})^{\mathbb{R}^{\mathbf{n}}}$ given as the set of $\mathcal{C}^\infty$-solutions of the system of linear constant-coefficient PDEs

$$R\left(\frac{\partial}{\partial v_1}, \frac{\partial}{\partial v_2}, \ldots, \frac{\partial}{\partial v_{\mathbf{n}}}\right) w = 0, \tag{37}$$

with $R \in \mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathbf{n}}]^{\bullet \times \mathbf{w}}$ a real polynomial matrix in $\mathbf{n}$ variables. The associated behavior is

$$\mathcal{B} = \left\{ w \in \mathcal{C}^\infty\left(\mathbb{R}^{\mathbf{n}}, \mathbb{R}^{\mathbf{w}}\right) \;\middle|\; R\left(\frac{\partial}{\partial v_1}, \frac{\partial}{\partial v_2}, \ldots, \frac{\partial}{\partial v_{\mathbf{n}}}\right) w = 0 \right\}.$$

The resulting system

$$\left(\mathbb{R}^{\mathbf{n}}, \mathbb{R}^{\mathbf{w}}, \mathtt{kernel}\left(R\left(\frac{\partial}{\partial v_1}, \frac{\partial}{\partial v_2}, \ldots, \frac{\partial}{\partial v_{\mathbf{n}}}\right)\right)\right)$$

is linear, since $w_1, w_2 \in \mathcal{B}$ and $\alpha, \beta \in \mathbb{R}$ imply $\alpha w_1 + \beta w_2 \in \mathcal{B}$, and shift invariant, since

$$w \in \mathcal{B} \quad \text{and} \quad (v_1, v_2, \ldots, v_{\mathbf{n}}) \in \mathbb{R}^{\mathbf{n}}$$

imply

$$\sigma^{(v_1, v_2, \ldots, v_{\mathbf{n}})} w \in \mathcal{B},$$

where

$$\begin{aligned} \left(\sigma^{(v_1, v_2, \ldots, v_{\mathbf{n}})} w\right)(v_1', v_2', \ldots, v_{\mathbf{n}}') \\ := w(v_1' + v_1, v_2' + v_2, \ldots, v_{\mathbf{n}}' + v_{\mathbf{n}}). \end{aligned}$$

Setting up these interconnection constraints constitutes linking.

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

Denote the set of linear shift-invariant nD differential systems or their behaviors by $\mathcal{L}_{\mathtt{n}}^{\mathtt{w}}$. The subscript $\mathtt{n}$ refers to the number of independent variables, while the superscript $\mathtt{w}$ is the number of dependent variables. When $\mathtt{n} > 1$, we refer to elements of the behavior as *fields*, while in the case $\mathtt{n} = 1$, these elements are *trajectories*. In the case of Maxwell's equations, $\mathtt{w} = 10$, and $\mathtt{n} = 4$, and the sparse, first order, polynomial matrix $R \in \mathbb{R}[\xi_t, \xi_x, \xi_y, \xi_z]^{8\times 10}$ corresponding to (37) with $w = \text{column}(E_x, E_y, E_z, B_x, B_y, B_z, j_x, j_y, j_z, \rho)$ is given by the equation shown at the bottom of the page.

Many, but by no means all, of the results obtained for 1D LTIDSs generalize to the nD case. In particular, the results discussed in "Three Central Theorems" generalize and require nothing more than an adaptation of the notation.

As in the 1D case, a behavior in $\mathcal{L}_{\mathtt{n}}^{\mathtt{w}}$ defines, through its linear constant-coefficient differential annihilators, an $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$-submodule of $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]^{\mathtt{w}}$. Conversely, since, as in the 1D-case, every $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$-submodule of $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]^{\mathtt{w}}$ is finitely generated, the generators of a submodule define a system (37), where the transposes of the rows of R are a set of generators of this submodule. This correspondence turns out to be one to one, but, for $\mathtt{n} > 1$, it is much more difficult to prove than in the 1D case that this correspondence is bijective.

The elimination theorem remains valid in the nD case. The projection of an element of $\mathcal{L}_{\mathtt{n}}^{\bullet}$ onto a subset of the dependent variables is again an element of $\mathcal{L}_{\mathtt{n}}^{\bullet}$. As an application, suppose that we want to characterize the behavior of the electrical variables $\vec{E}$, $\vec{j}$, and ρ in free space, without involving the magnetic field $\vec{B}$. Finding equations for the behavior of these electrical variables requires elimination of $\vec{B}$ from Maxwell's equations. The elimination theorem tells us that eliminating $\vec{B}$ leads to a system of linear constant-coefficient PDEs; this system is given by

A subset consisting of $\mathtt{m}$ of the $\mathtt{w}$ system variables is *free* in $\mathcal{B} \in \mathcal{L}_{\mathtt{n}}^{\mathtt{w}}$ if, after eliminating the $\mathtt{w} - \mathtt{m}$ remaining variables, the behavior of the $\mathtt{m}$ variables equals $\mathcal{C}^{\infty}(\mathbb{R}^{\mathtt{n}}, \mathbb{R}^{\mathtt{m}})$. For a minimal kernel representation in the 1D case, the number of system variables is equal to the sum of the number of equations (the number of rows of R) plus the maximal number of free variables. However, in the nD case, it is not always possible to obtain a kernel representation in which the number of system variables is equal to the sum of the number of equations plus the number of free variables. For example, in Maxwell's equations there are ten variables, eight equations, and three free variables, but it is not possible to write the behavior as the kernel of a system with only seven linear constant-coefficient PDEs.

One of the nice features of the behavioral notion of controllability is that it can be seamlessly adapted to nD systems (see [47] and [48] for discrete 2D systems and [34] for PDEs). Here, we discuss controllability of nD behaviors in $\mathcal{L}_{\mathtt{n}}^{\mathtt{w}}$. A behavior $\mathcal{B} \in \mathcal{L}_{\mathtt{n}}^{\mathtt{w}}$ is *controllable* if, for each pair of fields $w_1, w_2 \in \mathcal{B}$ and for each pair of open sets $O_1, O_2 \subset \mathbb{R}^{\mathtt{n}}$ with disjoint closures, there exists a field $w \in \mathcal{B}$ such that the restrictions of w_1 and w_2 satisfy $w|_{O_1} = w_1|_{O_1}$ and $w|_{O_2} = w_2|_{O_2}$. This definition is illustrated in Figure 14.

The question arises as to how to verify controllability of a system given by the kernel representation (37) and whether one can express controllability of nD systems in terms of a representation of the behavior. It turns out that, as for 1D systems, controllability for nD systems is equivalent to the existence of an image representation. Explicitly, $\mathcal{B} \in \mathcal{L}_{\mathtt{n}}^{\mathtt{w}}$ is controllable if and only if there exists a real polynomial matrix $M \in \mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]^{\mathtt{w}\times\bullet}$ such that $\mathcal{B} = \mathtt{image}\,(M(\partial/\partial v_1, \partial/\partial v_2, \ldots, \partial/\partial v_{\mathtt{n}}))$, with $M(\partial/\partial v_1, \partial/\partial v_2, \ldots, \partial/\partial v_{\mathtt{n}})$ viewed as an operator from

$$R(\xi_t, \xi_x, \xi_y, \xi_z) = \begin{bmatrix} \xi_x & \xi_y & \xi_z & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{\varepsilon_0} \\ 0 & -\xi_z & \xi_y & \xi_t & 0 & 0 & 0 & 0 & 0 & 0 \\ \xi_z & 0 & -\xi_x & 0 & \xi_t & 0 & 0 & 0 & 0 & 0 \\ -\xi_y & \xi_x & 0 & 0 & 0 & \xi_t & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \xi_x & \xi_y & \xi_z & 0 & 0 & 0 & 0 \\ -\xi_t & 0 & 0 & 0 & -c^2\xi_z & c^2\xi_y & -\frac{1}{\varepsilon_0} & 0 & 0 & 0 \\ 0 & -\xi_t & 0 & c^2\xi_z & 0 & -c^2\xi_x & 0 & -\frac{1}{\varepsilon_0} & 0 & 0 \\ 0 & 0 & -\xi_t & -c^2\xi_y & c^2\xi_x & 0 & 0 & 0 & -\frac{1}{\varepsilon_0} & 0 \end{bmatrix}.$$

By an interconnected system, we mean a system that consists of interacting subsystems modeled by tearing, zooming, and linking.

$\mathcal{C}^\infty(\mathbb{R}^{\mathtt{n}}, \mathbb{R}^\bullet)$ to $\mathcal{C}^\infty(\mathbb{R}^{\mathtt{n}}, \mathbb{R}^{\mathtt{w}})$. In fact, as in the 1D case, this existence of an image representation yields an algorithm for verifying controllability of a system $\mathcal{B} \in \mathcal{L}^{\mathtt{w}}_{\mathtt{n}}$ given in kernel representation. System (37) is controllable if and only if the $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$-module of annihilators is closed, equivalently, if and only if the syzygy of the right syzygy of R is equal to the $\mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]$-module spanned by the transposes of the rows of R (see "Polynomial Modules and Syzygies").

Let us consider the implications of this result for Maxwell's equations. These equations are usually presented as the kernel representation (37). However, there is an equivalent set of expressions that is useful for many derivations. In particular, it can be shown that the solutions of Maxwell's equations are exactly those that can be obtained by taking arbitrary potential fields $\vec{A} : \mathbb{R}^4 \to \mathbb{R}^3$ and $\phi : \mathbb{R}^4 \to \mathbb{R}$, and computing $\vec{E}, \vec{B}, \vec{j}, \rho$ through the equations

$$\begin{aligned}
\vec{E} &= -\frac{\partial}{\partial t}\vec{A} - \nabla\phi, \\
\vec{B} &= \nabla \times \vec{A}, \\
\vec{j} &= \varepsilon_0 \frac{\partial^2}{\partial t^2}\vec{A} - \varepsilon_0 c^2 \nabla^2 \vec{A} + \varepsilon_0 c^2 \nabla(\nabla \cdot \vec{A}) + \varepsilon_0 \frac{\partial}{\partial t}\nabla\phi, \\
\rho &= -\varepsilon_0 \frac{\partial}{\partial t}\nabla \cdot \vec{A} - \varepsilon_0 \nabla^2 \phi.
\end{aligned}$$

Since $\vec{A}$ and ϕ are free, these equations constitute an image representation of the behavior defined by Maxwell's equations. This representation of the solutions of Maxwell's equations as an image proves that Maxwell's equations define a controllable system. Representability of EM fields in free space as an image illustrates the importance of latent variables as well as of controllability. In (38), $\vec{A}$ and ϕ are latent variables introduced to display the solutions $\vec{E}, \vec{B}, \vec{j}, \rho$ of Maxwell's equations in a more convenient way. Maxwell's equations define a controllable system, and it is precisely the patchability of solutions expressed by controllability that makes this representation by means of potential fields possible.

One aspect in which the 1D case differs from the nD case is the existence of an observable image representation for a controllable system. The image representation $w = M(\partial/\partial v_1, \partial/\partial v_2, \ldots, \partial/\partial v_{\mathtt{n}})\ell$ is *observable* if there exists $N \in \mathbb{R}[\xi_1, \xi_2, \ldots, \xi_{\mathtt{n}}]^{\bullet \times \bullet}$ such that

$$w = M\left(\frac{\partial}{\partial v_1}, \frac{\partial}{\partial v_2}, \ldots, \frac{\partial}{\partial v_{\mathtt{n}}}\right)\ell$$

implies

$$\ell = N\left(\frac{\partial}{\partial v_1}, \frac{\partial}{\partial v_2}, \ldots, \frac{\partial}{\partial v_{\mathtt{n}}}\right)w.$$

In an observable image representation, the latent variables ℓ can be deduced from the manifest variables w and the laws that govern the system. In the 1D case, it is always possible to obtain an observable image representation for a controllable behavior. But controllable behaviors in $\mathcal{L}^{\mathtt{w}}_{\mathtt{n}}$ allow only exceptionally an observable image representation. In general the latent variables ℓ in an image representation are hidden, more precisely, not observable from the manifest variables. For example, Maxwell's equations do not allow an observable image representation. In particular, the underlying potentials $\vec{A}$ and ϕ cannot be deduced from the induced trajectory $\vec{E}, \vec{B}, \vec{j}, \rho$. The question of whether $\vec{A}$ and ϕ are physical quantities has been discussed in the physics literature [49]. But, in a sense, we need not worry about this lack of observability. The potential fields $\vec{A}$ and ϕ can simply be viewed as latent variables introduced to express the set of solutions $\vec{E}, \vec{B}, \vec{j}, \rho$ of Maxwell's equations in a more convenient way.

CONCLUSIONS

In this article we have presented an approach to the mathematical description of dynamical systems. The central notion is the behavior, which consists of the set of time trajectories that are declared possible by the model of a dynamical system. Often, the behavior is defined as the set of solutions of a system of differential equations. Models that specify a behavior usually involve latent variables in addition to the manifest variables the model aims at.

We have also described a methodology for modeling interconnected systems, called tearing, zooming, and linking. The underlying mathematical language consists of terminals, modules, the interconnection graph, the module embedding, and the manifest variable assignment. The combination of module equations, interconnection constraints, and manifest variable assignment leads to a latent-variable representation for the behavior of the manifest variables the model aims. This methodology of tearing, zooming, and linking offers a systematic procedure for modeling interconnected systems that is much better adapted to the physics of interconnected systems than input/output-modeling procedures such as, for example, Simulink. The methodology of tearing, zooming, and linking has many things in common with bond graphs.

In the tearing step, the system is viewed as an interconnection of subsystems.

The behavior is all there is. Modeling and system identification aim at a specification of the behavior. Properties and representations of systems refer to the behavior. We have illustrated this principle by the notions as linearity, time invariance, controllability, stabilizability, observability, and detectability.

In the later sections of this article, we discussed linear time-invariant differential systems. We concentrated on aspects of the theory of this class of systems that are new compared to systems defined by transfer functions or by state equations. Three key results are: i) linear time-invariant differential systems stand in one-to-one relation with $\mathbb{R}[\xi]$-modules, ii) complete elimination of latent variables is possible in the class of linear time-invariant differential equations, and iii) a linear time-invariant differential system is controllable if and only if its behavior admits an image representation. Each of these results generalizes verbatim to systems described by linear constant-coefficient PDEs. Another result is that a linear time-invariant differential input/output system is characterized by its transfer function if and only if it is controllable.

The main ideas of the behavioral point of view in modeling dynamical systems are summarized in "The Behavioral Approach." I believe that the main impact of this approach will eventually be felt on the level of elementary teaching.

Teaching System Theory

The behavioral approach offers an effective framework for teaching system theory on an elementary level. Focusing on the notion of a mathematical model, the behavior, as the first and central concept, allows for a balanced interplay between concrete physical examples and general mathematical ideas. Since no structure on the model equations is required, this approach can immediately deal with a broad class of examples. More refined system representations, with inputs, outputs, and states, can be brought in soon and positioned in the general setting. Mathematical concepts, such as Laplace transforms, transfer functions, and matrix exponentials, need not dominate the discussion from the beginning. Nothing more than simple set theory is required to introduce system properties such as controllability and observability. By thinking of control as restricting the plant behavior, one avoids the mathematical subtleties involved in explaining implicit equations that emerge in feedback structures.

Some remarks about teaching system theory from this vantage point can be found in "Teaching System Theory."

Throughout the 20th century, the field of systems and control was dominated by input/output thinking, where the interaction of a system with its environment is viewed as follows. The environment acts on the system by imposing an input, while the system reacts by imposing an output on the environment. This input/output view is eminently suitable in some cases, for example, in signal processing and in sensor-output-to-actuator-input feedback control. However, for modeling physical systems, it is often inappropriate. Input/output representations impose a cause/effect view of the interaction of a system with its environment, which is usually not part of the physical reality that the system describes. The difficulty with input/output thinking becomes even more pronounced in system interconnection. The requirement to endow a complex interconnected system with a signal flow graph to describe how subsystems interact is often arbitrary, and sometimes a caricature. This uneasiness with input/output thinking led to the development of the behavioral approach. A system is viewed as a family of trajectories, called the behavior. Interconnection is viewed as variable sharing. Control is viewed as restricting the plant behavior.

On the occasion of the 50th anniversary celebration of IFAC, on September 15, 2006, in Heidelberg, Karl Åström, expressed the view that "Block diagrams are unsuitable for serious physical modeling." He referred to this situation as the "control/physics barrier." Conventional systems and control thinking views interconnected systems in terms of block diagrams with input/output arrows connecting the building blocks, and signal-flow graphs showing the interconnection architecture. Indeed, this classical input/output view is unsuited for physical models since it forms a barrier to the application of system theory in modeling. The language of behaviors, as well as the methodology of tearing, zooming, and linking for modeling interconnected systems, overcomes this barrier. One can avoid arrows and signal flows, both in the basic dynamic models, as well as for interconnections.

In this article, I have examined the foundations of the field of systems and control. I have attempted to do so in a pedagogically responsible way, by confronting existing paradigms with concrete examples, and proposing new concepts when they are felt to be needed. Thinking of a dynamical system as a behavior, and of interconnection as variable sharing, gets the physics right. That is why I liked Kalman's premise mentioned at the beginning of this article.

ACKNOWLEDGMENTS

I thank Tzvetan Ivanov, Paolo Rapisarda, Gunther Reißig, and Shiva Shankar for commenting on some specific issues raised during the write-up of this article as well as numerous coworkers who helped me develop the ideas underlying the behavioral approach over the last two decades.

The SISTA-SMC research program is supported by the Research Council KUL: GOA AMBioRICS, CoE EF/05/006 Optimization in Engineering (OPTEC), IOF-SCORES4CHEM, several Ph.D./postdoc and fellow grants; by the Flemish Government: FWO: Ph.D./postdoc grants, projects G.0452.04 (new quantum algorithms), G.0499.04 (Statistics), G.0211.05 (Nonlinear), G.0226.06 (cooperative systems and optimization), G.0321.06 (Tensors), G.0302.07 (SVM/Kernel, research communities (ICCoS, ANMMM, MLDM); and IWT: Ph.D. Grants, McKnow-E, Eureka-Flite; by the Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007–2011) ; and by the EU: ERNSI.

REFERENCES

[1] J.C. Willems, "From time series to linear system—Part I. Finite dimensional linear time invariant systems, Part II. Exact modelling, Part III. Approximate modelling," *Automatica*, vol. 22, pp. 561–580, 675–694, 1986, and vol. 23, pp. 87–115, 1987.
[2] J.C. Willems, "Models for dynamics," *Dynamics Reported*, vol. 2, pp. 171–269, 1989.
[3] J.C. Willems, "Paradigms and puzzles in the theory of dynamical systems," *IEEE Trans. Automat. Contr.*, vol. 36, no. 3, pp. 259–294, 1991.
[4] J.W. Polderman and J.C. Willems, *Introduction to Mathematical Systems Theory: A Behavioral Approach*. New York: Springer-Verlag, 1998.
[5] J.C. Willems, "System theoretic models for the analysis of physical systems," *Ricerche di Automatica*, vol. 10, pp. 71–106, 1979.
[6] V. Belevich, *Classical Network Theory*. San Francisco, CA: Holden-Day, 1968.
[7] R.W. Newcomb, *Linear Multiport Synthesis*. New York: McGraw-Hill, 1966.
[8] G.J. Klir, *An Approach to General Systems Theory*. New York: Van Nostrand Reinhold, 1969.
[9] T.G. Windeknecht, *General Dynamical Processes*. New York: Academic Press, 1971.
[10] S. Eilenberg Automata, *Languages, and Machines*. New York: Academic Press, New York, 1974.
[11] D. Cox, J. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms*. New York: Springer-Verlag, 1997.
[12] M. Fliess, J. Lévine, Ph. Martin, and P. Rouchon, "Flatness and defect of nonlinear systems: Introductory theory and applications," *Int. J. Contr.*, vol. 61, pp. 1327–1369, 1995.
[13] L.D. Chua and R.A. Rohrer, "On the dynamic equations of a class of nonlinear RLC networks," *IEEE Trans. Circuits Theory*, vol. 12, pp. 475–489, 1965.
[14] W.M. Haddad, V. Chellaboina, and S. Nersesov, *Thermodynamics, A Dynamical Systems Approach*. Princeton, NJ: Princeton Univ. Press, 2005.
[15] J.C. Willems, "Review of the book Thermodynamics: A Dynamical Systems Approach, by W.M. Haddad, V.S. Chellaboina, and S. Nersesnov," *IEEE Trans. Automat. Contr.*, vol. 51, pp. 1217–1225, 2006.
[16] T. Cotroneo, "Algorithms in behavioral systems theory," Ph.D. dissertation, Univ. Groningen, 2001. [Online]. Available: http://dissertations.ub.rug.nl/faculties/science/2001/t.cotroneo
[17] T. Cotroneo and J.C. Willems, "The simulation problem for high order differential equations," *Appl. Math. Comput.*, vol. 145, pp. 821–851, 2003.
[18] PSpice Tutorials [Online]. Available: http://denethor.wlu.ca/PSpice and http://dave.uta.edu/dillon/pspice/index.php
[19] Modelica and the Modelica Assoc. [Online]. Available: http://www.modelica.org
[20] The Mathworks, Simulink [Online]. Available: http://www.mathworks.com/products/simulink
[21] Bond graphs [Online]. Available: http://www.20sim.com/product/bondgraphs.html.
[22] P.J. Gawthrop and G.P. Bevan, "Bond-graph modeling," *IEEE Control Syst. Mag.*, vol. 27, no. 2, pp. 24–45, 2007.
[23] A.J. van der Schaft, "Port-Hamiltonian systems," in *Modeling and Control of Complex Physical Systems—The Port-Hamiltonian Approach*, to appear.
[24] P. Rapisarda, "Linear differential systems," Ph.D. dissertation, Univ. Groningen, 1998. [Online]. Available: http://dissertations.ub.rug.nl/faculties/science/1998/p.rapisarda}
[25] P. Rapisarda and J.C. Willems, "State maps for linear systems," *SIAM J. Control Optim.*, vol. 35, pp. 1053–1091, 1997.
[26] M. Kuijper, *First-Order Representations of Linear Systems*. Cambridge, MA: Birkhäuser, 1994.
[27] P.A. Fuhrmann, P. Rapisarda, and Y. Yamamoto, "On the state of behaviors," *Linear Algebra Its Applications*, vol. 424, pp. 570–614, 2007.
[28] R.E. Kalman, "Contributions to the theory of optimal control," *Boletin de la Sociedad Matematica Mexicana*, vol. 5, pp. 102–119, 1960.
[29] O. Mayr, *The Origins of Feedback Control*. Cambridge, MA: MIT Press, 1970.
[30] M.C. Smith, "Synthesis of mechanical networks: The inerter," *IEEE Trans. Automat. Contr.*, vol. 47, no. 10, pp. 1648–1662, 2002.
[31] S. Evangelou, D.J.N. Limebeer, R.S. Sharp, and M.C. Smith, "Control of motorcycle steering instabilities—Passive mechanical compensators incorporating inerters," *IEEE Control Syst. Mag.*, vol. 26, no. 5, pp. 78–88, 2006.
[32] P.A. Fuhrmann, "A study of behaviors," *Linear Algebra Its Applications*, vol. 351–352, pp. 303–380, 2002.
[33] F. Fagnani and J.C. Willems, "Representations of symmetric linear dynamical systems," *SIAM J. Control Optim.*, vol. 31, pp. 1267–1293, 1993.
[34] H.K. Pillai and S. Shankar, "A behavioral approach to control of distributed systems," *SIAM J. Control Optim.*, vol. 37, pp. 388–408, 1999.
[35] H. Glüsing-Lüerssen, "A behavioral approach to delay-differential systems," *SIAM J. Control Optim.*, vol. 35, pp. 480–499, 1997.
[36] P. Rocha and J.C. Willems, "Behavioral controllability of delay-differential systems," *SIAM J. Control Optim.*, vol. 35, pp. 254–264, 1997.
[37] J.C. Willems and Y. Yamamoto, "Behaviors defined by rational functions," *Linear Algebra and Its Applications*, vol. 425, pp. 226–241, 2007.
[38] D.G. Meyer, "Fractional balanced reduction: Model reduction via fractional representation," *IEEE Trans. Automat. Contr.*, vol. 35, no. 12, pp. 1341–1345, 1990.
[39] J.C. Willems, "On interconnections, control, and feedback," *IEEE Trans. Automat. Contr.*, vol. 42, no. 3, pp. 326–339, 1997.
[40] M.N. Belur and H.L. Trentelman, "On stabilization, pole placement and regular implementability," *IEEE Trans. Automat. Contr.*, vol. 47, no. 5, pp. 735–744, 2002.
[41] M.N. Belur, "Control in a behavioral context," Ph.D. dissertation, Univ. Groningen, 2003. [Online]. Available: http://irs.ub.rug.nl/ppn/250129701.
[42] R.F. Curtain and H. Zwart, *An Introduction to Infinite-Dimensional Linear Systems Theory*. New York: Springer-Verlag, 1995.
[43] U. Oberst, "Multidimensional constant linear systems," *Acta Applicandae Mathematicae*, vol. 20, pp. 1–175, 1990.
[44] E. Zerz, *Topics in Multidimensional Linear System Theory*, vol. 256. New York: Springer Verlag Lecture Notes in Control and Information Sciences, 2000.
[45] J.F. Pommaret, *Partial Differential Equations and Group Theory: New Perspectives for Applications*, Norwell, MA: Kluwer, 1999.
[46] J.F. Pommaret and A. Quadrat, "Algebraic analysis of linear multidimensional control systems," *IMA J. Math. Contr. Inform.*, vol. 16, pp. 275–297, 1999.
[47] P.M. Rocha, "Structure and represenation of 2-D systems," Ph.D. dissertation, Univ. Groningen, 1990.
[48] P.M. Rocha and J.C. Willems, "Controllability of 2-D systems," *IEEE Trans. Automat. Contr.*, vol. 36, no. 4, pp. 413–423, 1991.
[49] E. Scheibe, "On the mathematical overdetermination of physics," in *Philosophy, Mathematics and Modern Physics*, E. Rudolph and I.-O. Stamatescu, Eds. New York: Springer-Verlag, pp. 186–199, 1994.

AUTHOR INFORMATION

Jan Willems (jan.willems@esat.kuleuven.be) graduated from the University of Gent in 1963 with a combined degree in electrical and mechanical engineering. He obtained an M.Sc. degree from the University of Rhode Island and a Ph.D. degree from MIT in 1968. He was on the electrical engineering faculty of MIT from 1968 to 1973, including a one-year postdoctoral at Cambridge University. In 1973 he became a professor of systems and control with the Mathematics Department of University of Groningen. In 2003, he became emeritus. He is presently a guest professor at the K.U. Leuven. Details of his research activities can be found on his Web site http://homes.esat.kuleuven.be/~jwillems/webpage.html.