Input-Output and State-Space Representations of Finite-Dimensional Linear Time-Invariant Systems

Jan C. Willems
Mathematics Institute
P. O. Box 800
9700 AV Groningen, The Netherlands

Submitted by Paul A. Fuhrmann

ABSTRACT

In this paper we develop various representations for systems described by a set of high-order differential equations of the form $R_0w + R_1\dot{w} + \cdots + R_s w^{(s)} = 0$, with $R_0, R_1, \dots, R_s$ not necessarily square matrices. The variables w are the external variables. Particular attention is paid to the problem of obtaining minimal state-space realizations and input-output or input-state-output representations of such systems.

1. INTRODUCTION

It is customary in systems theory to deal with mathematical models of dynamical systems which are driven by inputs and which produce outputs. This is the common starting point both in the external "input-output" framework and in the internal "state-space" framework. Thus one assumes in effect that it can be postulated from general considerations (the signal flow graph of the system) through which variables the environment influences the system and through which variables the system influences its environment in turn. However, there are many situations (for a number of examples, see [1]) where such a cause-effect relation is *not* a natural starting point and in which an input-output model appears as a specific structure-oriented *representation* of a system. Consequently there is a need to develop a framework in which it can be decided on the basis of the mathematical model (or, more generally, on the basis of the data) what are the inputs and outputs. Of course, in order for this to make sense we need to start from a more general vantage point for the description of dynamical systems, a starting point in which the input-output distinction is not made *ab initio*. In [1] we have presented such a framework. In the present paper we will take a closer look at *continuous-time linear*

52 Vanderbilt Ave., New York, NY 10017 0024-3795/83/020581-28$3.00

time-invariant finite-dimensional systems (some preliminary results were announced in [2]). We start by briefly describing some examples justifying our axiomatic starting point. More details and other examples may be found in [1].

2. MOTIVATIONAL EXAMPLES

EXAMPLE 2.1 (Kepler's laws—autonomous systems). According to Kepler's laws, planets move in elliptical orbits with the sun at a focus, such that the radius vector to the sun sweeps out equal areas in equal times, and such that the ratio of the square of the period of revolution to the cube of the major axis of the ellipse is a constant. The collection of all such orbits constitutes a well-defined family of trajectories. In what sense is it a dynamical system? What is its state? We know of course that we can obtain these motions as the solutions of a Hamiltonian system of differential equations. This system is however an autonomous one: there are no external inputs, and hence its state-space realization is not covered by the input-output approach to realization theory. We shall see that autonomous systems fit very naturally in our framework (for more details on the realization of Kepler's laws via Newton's equations, see [1]).

The absence of externally defined autonomous systems is a bit of an annoying drawback of the classical input-output Nerode-equivalence–Hankel-matrix approach to realization theory. This drawback is absent in our framework. (We should mention however that also in Fuhrmann's "polynomial model" realization algorithm, nonreachable modes are not necessarily canceled out). It is not true for example that one can identify only the reachable and observable part of a system (notwithstanding many statements in the literature to the contrary). As a trivial example, the estimation of the trend parameter α in the autonomous system $y(t+1)=\alpha y(t)$ is a natural and important question to ask in the identification of linear systems.

EXAMPLE 2.2. Consider an electrical circuit consisting of the interconnection of a finite number of resistors, inductors, capacitors, transformers, and gyrators, and with a number of external driving ports. We can easily write down the equations which its branch currents and voltages have to obey. These consist of Kirchhoff's current and voltage laws and of the differential equations (for the L's and C's) or the algebraic equations (for the R's, T's and G's) which express the constraints imposed by the constitutive laws of the elements appearing in the branches. How should we define the external-port behavior of this circuit? It need not have an admittance or an impedance representation. Indeed, which port variables can be considered as inputs and outputs depends on the circuit configuration. Nevertheless, the model is completely specified, we know (in principle) which external signals can occur, and the voltages on the capacitors and the currents through the inductors should qualify as state variables. We shall see that in our framework we can treat this situation very nicely (see Example 3.2).

3. SYSTEMS IN EXTERNAL FORM

We shall consider in this paper only continuous-time systems with time axis $T = \mathbb{R}$. A *dynamical system* Σ is then defined as a subset of $W^{\mathbb{R}}$, with W a set called the *external signal alphabet*. Thus a system is simply a collection of maps from $\mathbb{R}$ into W, i.e., a family of trajectories. Σ is said to be *time-invariant* if $\forall \tau \in \mathbb{R}$ we have $S_\tau \Sigma = \Sigma$, where S_τ denotes the τ-shift, i.e., $S_\tau f : \mathbb{R} \to W$ is defined by $(S_\tau f)(t) := f(t - \tau)$. Σ is said to be *linear* if W is a vector space and Σ is a linear subspace of $W^{\mathbb{R}}$. We shall denote vector spaces by script capitals in the sequel.

Let $\mathbb{R}[s]$ denote, as usual, the real polynomials in the indeterminate s, $\mathbb{R}(s)$ the rational functions, $\mathbb{R}^n[s]$ the n-dimensional vectors of real polynomials, $\mathbb{R}^{n_1 \times n_2}[s]$ the $n_1 \times n_2$ matrices of real polynomials, etc. An element of $\mathbb{R}(s)$ is said to be *(strictly) proper* if the degree of its denominator is (strictly) larger than that of its numerator. Similarly for vectors and matrices of rational functions.

The special class of systems which we will study in detail in this paper is defined by a real polynomial matrix $R \in \mathbb{R}^{r \times q}[s]$, with $\mathcal{W} = \mathbb{R}^q$, as follows:

$$\Sigma(R) := \left\{ w : \mathbb{R} \to \mathcal{W} \,|\, w \in \mathfrak{L}^{\mathrm{loc}} \text{ and } R\left(\frac{d}{dt}\right) w = 0 \right\}$$

Here $\mathfrak{L}^{\mathrm{loc}}$ denotes the locally integrable vector-valued functions on $\mathbb{R}$, and $R(d/dt)w = 0$ is to be interpreted in the sense of distributions. Obviously $\Sigma(R)$ is linear and time-invariant.

EXAMPLE 3.1. Let $P \in \mathbb{R}^{p \times p}[s]$, $Q \in \mathbb{R}^{p \times m}[s]$, $\det P(s)$ be unequal to the zero polynomial, and $P^{-1}(s)Q(s)$ be a proper rational matrix. Consider now the set of differential equations

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

This is obviously a system of the type $\Sigma(R)$ with

$$\mathcal{W} = \mathbb{R}^{m+p}, \qquad w = \begin{bmatrix} u \\ y \end{bmatrix}, \quad \text{and} \quad R = [\,Q \,\vdots\, -P\,].$$

It has a special form because the first m components deserve to be called *inputs* and the others *outputs*. Its *transfer function* matrix is $P^{-1}Q$. Note, however, that also responses not explainable by inputs but entirely due to initial conditions may occur in $\Sigma([Q \,\vdots\, -P])$. In fact the autonomous case with $Q = 0$ is of particular interest.

Note that the elements of $\Sigma(R)$ need not have support on a half line of the type $[a, \infty)$. In the classical input-output framework it is customary to assume such a half-line support. This difference is an important one. It is motivated as follows. Most of all, we feel that our definition is more natural. It considers the differential equations as the basic description modeling the behavior, without introducing assumptions which may be motivated by mathematical expediency. In addition, it is essential not to assume this half-line support if one wants to incorporate autonomous systems.

Observe that we have postulated no smoothness in $\Sigma(R)$, other than $w \in \mathcal{L}^{\text{loc}}$. This is nothing unusual. It is in fact what is common in control theory, as shown in Example 3.1, where (disregarding initial conditions) the map $u \mapsto y$ is a smooth convolution and hence we obtain $y \in \mathcal{L}^{\text{loc}}$ if $u \in \mathcal{L}^{\text{loc}}$. We shall return to this smoothness issue in Section 8.3.

Often systems are defined in terms of auxiliary variables. A general such class may be defined by $R_1 \in \mathbb{R}^{r \times q}[s]$ and $R_2 \in \mathbb{R}^{r \times t}[s]$, with $\mathcal{W} = \mathbb{R}^q$, as follows:

$$\Sigma(R_1, R_2) := \Big\{ w : \mathbb{R} \to \mathcal{W} \,|\, w \in \mathcal{L}^{\text{loc}} \text{ and there is an } \mathbb{R}^t\text{-vector-valued distribution } \xi \text{ such that } R_1\left(\frac{d}{dt}\right)w = R_2\left(\frac{d}{dt}\right)\xi \Big\}$$

with, again, equality in the sense of distributions. Clearly $\Sigma(R_1, R_2)$ is also linear and time-invariant.

An example of such a class of systems which will play an important role later in our paper is state-space systems. Another example is the systems studied by Rosenbrock [4] and Wolovich [5] (see also [6]). There one starts with polynomial matrices P, Q, R, W and looks at $P(d/dt)\xi = Q(d/dt)u$, $y = R(d/dt)\xi + W(d/dt)u$, with ξ what is called the *partial state*. Writing

$$w = \begin{bmatrix} u \\ y \end{bmatrix}$$

yields a system of the type $\Sigma(R_1, R_2)$. (As we shall explain later, we do not call u the input and y the output unless $W + RP^{-1}Q$ is proper.)

EXAMPLE 3.2. An example of a physical system defined in terms of auxiliary variables is an electrical circuit. Indeed, it is most natural to view the port behavior of an *RLCTG* network as described in Example 2.2 as follows. Let (V_e, I_e) denote the vector of *external* port voltages and currents, and (V_i, I_i) the vector of *internal* branch voltages and currents. These will satisfy the equations $K_v(V_i, I_i, V_e, I_e) = 0$, $K_c(V_i, I_i, V_e, I_e) = 0$, $L(V_i, \dot{I}_i) = 0$, $C(\dot{V}_i, I_i) = 0$, $R(V_i, I_i) = 0$, $T(V_i, I_i) = 0$, and $G(V_i, I_i) = 0$, where K_v, K_c, L, C, R, T, and G denote respectively the equations obtained by writing out Kirchhoff's voltage and current laws and the constitutive equations of the inductors, capacitors, resistors, transformers, and gyrators. Since these equations are linear, this clearly leads to a set of equations as in $\Sigma(R_1, R_2)$, with

$$w = \begin{bmatrix} V_e \\ I_e \end{bmatrix} \quad \text{and} \quad \xi = \begin{bmatrix} V_i \\ I_i \end{bmatrix}.$$

Any system described with auxiliary variables can also be described without. Indeed:

PROPOSITION 3.3. *Let $\Sigma(R_1, R_2)$ be given. Then there exists R such that $\Sigma(R) = \Sigma(R_1, R_2)$.*

Proof. Clearly for any unimodular matrices M_1 and M_2 we have $\Sigma(R_1, R_2) = \Sigma(M_1R_1, M_1R_2M_2)$. Now choose M_1 and M_2 to bring $M_1R_2M_2 =: R_3$ into Smith form [6, p. 390]:

$$R_3(s) = \begin{bmatrix} \operatorname{diag}(p_1(s), \ldots, p_k(s)) & 0 \\ 0 & 0 \end{bmatrix}$$

with $0 \neq p_i \in \mathbb{R}[s]$, $i = 1, \ldots, k$. Now partition M_1R_1 conformably as $\begin{bmatrix} R_{11} \\ R_{12} \end{bmatrix}$.

The system $\Sigma(R_1, R_2)$ is hence given by

$$\Sigma(R_1, R_2) = \Big\{ w : \mathbb{R} \to \mathcal{W} \,|\, w \in \mathcal{L}^{\text{loc}}$$

and there are distributions ξ_i such that

$$R_{11}\left(\frac{d}{dt}\right)w = \text{col}\left(p_1\left(\frac{d}{dt}\right)\xi_1, \dots, p_k\left(\frac{d}{dt}\right)\xi_k\right)$$

$$\text{and } R_{12}\left(\frac{d}{dt}\right)w = 0\Big\}.$$

However, *any* real-valued distribution can be written as $p_i(d/dt)\xi_i$ for some distribution ξ_i. Consequently the first set of equations in this expression imposes no restriction on w. Hence $\Sigma(R_1, R_2) = \Sigma(R_{12})$. ∎

The systems usually considered in linear systems theory are those defined in Example 3.1. The question hence occurs whether any system can be brought in this form. This is indeed the case.

DEFINITION 3.4. Let $\Sigma(R)$ be given. If T is a constant nonsingular matrix such that, with P, Q as in Example 3.1, $\Sigma(RT^{-1}) = \Sigma([Q \,|\, -P])$, then we call $\{(P(s), Q(s)), T\}$ an *input-output* (i/o) *representation* of $\Sigma(R)$, with the first m components of Tw the *inputs* and the last p components the *outputs*. Clearly $m + p = r$.

THEOREM 3.5. *Let* $\Sigma(R)$ be given. Then it admits an i/o representation $\{(P(s), Q(s)), T\}$. In fact, T may be chosen to be a permutation matrix. Alternatively, we may choose T such that $P^{-1}Q$ is strictly proper.

Proof. For any unimodular matrix M we have that $\Sigma(MR) = \Sigma(R)$. It is well known (see e.g. [4, p. 30]) that M can always be chosen such that MR is row proper, i.e. such that

$$M(s)R(s) = \text{diag}(s^{n_1}, \dots, s^{n_k})\begin{bmatrix} P_0 \\ 0 \end{bmatrix}$$
$$+ \text{diag}(s^{n_1 - 1}, \dots, s^{n_k - 1})\begin{bmatrix} P_1 \\ 0 \end{bmatrix} + \cdots$$

(where negative powers of s would appear, read zero). By choosing T suitably we obtain

$$M(s)R(s)T^{-1} = \mathrm{diag}(s^{n_1},\dots,s^{n_k})\begin{bmatrix} P_0' & P_0'' \\ 0 & 0 \end{bmatrix}$$
$$+\,\mathrm{diag}(s^{n_1-1},\dots,s^{n_k-1})\begin{bmatrix} P_1' & P_1'' \\ 0 & 0 \end{bmatrix} + \cdots$$

with P_0' square and nonsingular. Obviously T^{-1} may be chosen as the permutation matrix which selects independent columns from P_0. Alternatively T^{-1} can be chosen such that $P_0'' = 0$. Now by defining $P(s) := \mathrm{diag}(s^{n_1},\dots,s^{n_k})P_0' + \mathrm{diag}(s^{n_1},\dots,s^{n_k-1})P_1' + \cdots$, and $Q(s) := \mathrm{diag}(s^{n_1},\dots,s^{n_k})P_0'' + \mathrm{diag}(s^{n_1-1},\dots,s^{n_k-1})P_1'' + \cdots$, we obtain an i/o representation. Indeed, $w \in \Sigma(R)$ iff $Tw \in \Sigma(RT^{-1}) = \Sigma(MRT^{-1}) = \Sigma([Q \mid -P])$. Furthermore, since $\lim_{s\to\infty} P^{-1}(s)Q(s) = (P_0')^{-1}P_0''$, $P^{-1}Q$ is proper, as required. By choosing T^{-1} as mentioned, we obtain an i/o representation with T a permutation matrix, or with $P^{-1}Q$ strictly proper. ■

We shall comment on the significance of this theorem in Remark 6.4.

4. SYSTEMS IN STATE-SPACE FORM

A *dynamical system in state-space form* is defined as a subset $\Sigma_i \subset (X \times W)^{\mathbb{R}}$, with X the *state space* and W the *external signal alphabet*. It needs to satisfy the following axiom, which formalizes, in a set-theoretic sense, that *the past and the future behavior are independent given the present state*. Let $t \in \mathbb{R}$ and $a \in X$, and consider the following sets:

$$\Sigma_{i,t}^-(a) := \{(x^-, w^-) \mid :(-\infty, t) \to X \times W \mid \exists (x, w) \in \Sigma_i \ni x(t) = a$$
$$\text{and } (x^-, w^-)(\tau) = (x, w)(\tau) \text{ for } \tau < t\},$$

$$\Sigma_{i,t}^+(a) := \{(x^+, w^+) \mid :[t, \infty) \to X \times W \mid \exists (x, w) \in \Sigma_i \ni x(t) = a$$
$$\text{and } (x^+, w^+)(\tau) = (x, w)(\tau) \text{ for } \tau \geqslant t\},$$

$$\Sigma_{i,t}(a) := \{(x, w) \mid : \mathbb{R} \to X \times W \mid (x, w) \in \Sigma_i \text{ and } x(t) = a\}.$$

The *axiom of state* demands that for all $t \in \mathbb{R}$ and $a \in X$ there holds

$\Sigma_{i,t}(a)=\Sigma_{i,t}^{-}(a)\cdot\Sigma_{i,t}^{+}(a)$, where $\cdot$ denotes the concatenation product. The definition of *linearity* and *time invariance* carry over unchanged from the previous section.

We now define the class of state-space systems which we will study. Consider the system $\Sigma_i(A, B, C, D)$: $\dot{x}=Ax+Bu$, $w=Cx+Du$ with $x\in\mathcal{X}=\mathbb{R}^n$, $u\in\mathcal{U}=\mathbb{R}^m$, and $w\in\mathcal{W}=\mathbb{R}^q$. Formally $\Sigma_i(A, B, C, D):=\{(x,w):\mathbb{R}\to\mathcal{X}\times\mathcal{W}\mid x$ is absolutely continuous, $u\in\mathfrak{L}^{\text{loc}}$, $\dot{x}(t)=Ax(t)+Bu(t)$ for almost all t, and $w(t)=Cx(t)+Du(t)$ for all $t\}$. It is easy to see that this defines a linear time-invariant state-space system. A somewhat more general class of such systems is given by the (possibly singular) set of first-order differential equations $E\dot{x}=Ax+Bw$.

For a given state-space system Σ_i we define its *external behavior* as $\Sigma_e:=\{w:\mathbb{R}\to W\mid\exists x\ni(x,w)\in\Sigma_i\}$. Obviously the external behavior of a linear and/or time-invariant system is linear and/or time-invariant. If Σ is a given externally defined dynamical system which equals the external behavior of Σ_i, then we shall say that Σ_i is a (state-space) *realization* or a *state-space representation* of Σ. From now on we shall assume all the systems to be time-invariant.

Recall the definitions of $\Sigma_{i,0}(a)$, $\Sigma_{i,0}^{-}(a)$, and $\Sigma_{i,0}^{+}(a)$ as given in the beginning of this section. Now define analogously the external versions of these objects:

$$\Sigma_e(a):=\{w:\mathbb{R}\to W\mid\exists x\ni(x,w)\in\Sigma_{i,0}(a)\},$$

$$\Sigma_e^{-}(a):=\{w^{-}:(-\infty,0)\to W\mid\exists x^{-}\ni(x^{-},w^{-})\in\Sigma_{i,0}^{-}(a)\},$$

$$\Sigma_e^{+}(a):=\{w^{+}:[0,\infty)\to W\mid\exists x^{+}\ni(x^{+},w^{+})\in\Sigma_{i,0}^{+}(a)\}.$$

Obviously $\Sigma_e(a)\subset\Sigma_e^{-}(a)\cdot\Sigma_e^{+}(a)$. By the axiom of state we actually have $\Sigma_e(a)=\Sigma_e^{-}(a)\cdot\Sigma_e^{+}(a)$. Also $\Sigma_e=\cup_{a\in X}\Sigma_e(a)=\cup_{a\in X}\Sigma_e^{-}(a)\cdot\Sigma_e^{+}(a)$. In other words, if we regard Σ_e as a relation on (i.e., a subset of) $W^{(-\infty,0)}\times W^{[0,\infty)}$, then a state-space realization simply induces a partition of this relation into the join of product relations (i.e., into a union of rectangular subsets $\Sigma_e^{-}(a)\cdot\Sigma_e^{+}(a)$ of $W^{(-\infty,0)}\times W^{[0,\infty)}$). It is in this context that we define the minimality of a state-space system viewed as a realization of its own external behavior. Thus Σ_i is said to be *minimal* if:

(1) Whenever $\cup_{a\in X'}\Sigma_e(a)=\Sigma_e$, then we must have that $X'=X$. (This says that none of the $\Sigma_e(a)$'s will be empty and that none of them will be covered by the others. Indeed, if that were the case, we could *delete* this state from the state space and obtain a reduced realization.)

(2) Whenever $\bigcup_{a \in X'} \Sigma_e(a)$ is rectangular (i.e., it can be written as $R^- \cdot R^+$ with $R^- \subset W^{(-\infty,0)}$ and $R^+ \subset W^{(0,\infty)}$), then X' must consist of at most one point (for otherwise, by *combining* the states in X' into one, we would obtain a reduced realization).

Two state-space systems $\Sigma_i^1 \subset (X_1 \times W)^{\mathbb{R}}$ and $\Sigma_i^2 \subset (X_2 \times W)^{\mathbb{R}}$ are said to be *equivalent* if there exists a bijection $S: X_1 \to X_2$ such that $\{(x_1, w) \in \Sigma_i^1\} \Leftrightarrow \{(Sx_1, w) \in \Sigma_i^2\}$. Obviously two equivalent systems have the same external behavior, and a natural question to ask is if all minimal realizations of a given externally defined system are equivalent. This issue is studied in [1]. In general the answer is *no* (contrary to what happens in the classical input-output setting). There is however a natural condition for all minimal realizations to be equivalent. Indeed, in [1] it is shown that all minimal realizations of Σ are equivalent iff whenever two pasts (futures) of trajectories of Σ have one future (past) in common, then they have *all* their futures (pasts) in common. Let us denote by f^+ (f^-) the restriction of a map on $\mathbb{R}$ to $(-\infty,0)$ ($[0,\infty)$). Formally:

Proposition 4.1. *All minimal realizations of an externally defined time-invariant system Σ are equivalent iff*

(i) $\{w_1^- \cdot w^+, w_2^- \cdot w^+ \in \Sigma\} \Rightarrow \{\{w_1^- \cdot v^+ \in \Sigma\} \Leftrightarrow \{w_2^- \cdot v^+ \in \Sigma\}\}$ *and*
(ii) $\{w^- \cdot w_1^+, w^- \cdot w_2^+ \in \Sigma\} \Rightarrow \{\{v^- \cdot w_1^+ \in \Sigma\} \Leftrightarrow \{v^- \cdot w_2^+ \in \Sigma\}\}$

In terms of the notation introduced before, the above proposition states that all minimal realizations of Σ are equivalent iff Σ may be written as

$$\Sigma = \bigcup_{\alpha \in A} R_\alpha^- \cdot R_\alpha^+ \text{ with in addition } \{\alpha', \alpha'' \in A, \alpha' \neq \alpha''\}$$

$$\Rightarrow \quad \{R_{\alpha'}^- \cap R_{\alpha''}^- = \varnothing \text{ and } R_{\alpha'}^+ \cap R_{\alpha''}^+ = \varnothing\}.$$

From the above proposition it follows that all minimal realizations of a *linear* time-invariant system are equivalent. (The reason for introducing in the present paper the more general set-theoretic version of minimality was precisely the fact that we can obtain the equivalence of all minimal realizations as a result which uses linearity in an essential way.) Indeed, let $\Sigma^+(0)$ be defined as

$$\Sigma^+(0) := \{ w^+ : [0,\infty) \to W | 0^- \cdot w^+ \in \Sigma\},$$

and let $w^- \cdot w^+ \in \Sigma$. Then $\{w^- \cdot w^+ \in \Sigma\} \Rightarrow \{\{w^- \cdot w^+ \in \Sigma\} \Leftrightarrow \{v^+ \in w^+ +$

$\Sigma^+(0))\}$. Consequently $\{w_1^- \cdot w^+, w_2^- \cdot w^+ \in \Sigma\} \Rightarrow \{\{w_1^- \cdot v^+ \in \Sigma\} \Leftrightarrow \{v^+ \in w^+ + \Sigma^+(0)\} \Leftrightarrow \{w_2^- \cdot v^+ \in \Sigma\}\}$, which by Proposition 4.1 yields that all minimal realizations of a linear time-invariant system are equivalent. The following theorem shows, among other things, that for this class of systems minimality is equivalent to the state space having as small as possible a dimension.

THEOREM 4.2.

(1) *All minimal realizations of a linear time-invariant system are equivalent. In fact, if* $\Sigma^1_{\min}$ *is a minimal linear time-invariant realization of* Σ, *then so is* $\Sigma^2_{\min}$ *iff there exists a linear bijection* S *(from the state space of* $\Sigma^1_{\min}$ *onto that of* $\Sigma^2_{\min}$*) such that* $\{(x,w)\in\Sigma^1_{\min}\} \Leftrightarrow \{(Sx,w)\in\Sigma^2_{\min}\}$. *In fact, if* $\Sigma^1_{\min}$ *and* $\Sigma^2_{\min}$ *are both minimal linear time-invariant systems and* S *is a bijection demonstrating their equivalence, then* S *is linear.*

(2) *Let* Σ_i *and* $\Sigma_{\min}$ *be linear time-invariant systems with state spaces* X *and* $X_{\min}$ *respectively, which both realize* Σ, *and let* $\Sigma_{\min}$ *be minimal. Then there exists a linear subspace* X' *of* X *and a linear surjective map* $S: X' \to X_{\min}$ *such that* $\{(x,w)\in\Sigma_i\} \Leftrightarrow \{x(t)\in X'$ *for all* t *and* $(Sx,w)\in\Sigma_{\min}\}$. *Consequently* $\dim X_{\min} \leqslant \dim X$.

(3) *Finally, if* Σ *has a linear time-invariant realization with a finite-dimensional state space, then minimality is equivalent to its state space having as small as possible a dimension in the class of linear time-invariant realizations.*

We shall not prove this theorem, since it is not particularly germane to the rest of the paper. It follows without much difficulty from the ideas in [1].

With these general definitions and results in mind regarding general linear time-invariant systems, we now return to $\Sigma(R)$ and $\Sigma_i(A,B,C,D)$. The external behavior of $\Sigma_i(A,B,C,D)$ will be denoted by $\Sigma_e(A,B,C,D)$. Note that $\Sigma_e(A,B,C,D)$ is not *a priori* equal to the system which we would obtain by letting x and u in $\dot{x} = Ax + Bu$, $w = Cx + Du$ be auxiliary variables as in the definition of $\Sigma(R_1,R_2)$. This because x and u could then be distributions, while in $\Sigma_e(A,B,C,D)$ it is assumed that x is absolutely continuous. We can now ask the following questions:

(1) When is $\Sigma_i(A,B,C,D)$ a minimal realization of its own external behavior $\Sigma_e(A,B,C,D)$?

(2) When is $\Sigma_i(A,B,C,D)$ a minimal realization of the system obtained by considering x and u as auxiliary variables? We shall denote this system by $\Sigma_e'(A,B,C,D)$.

For emphasis, we repeat the definitions of $\Sigma_e(A, B, C, D)$ and $\Sigma'_e(A, B, C, D)$:

$$\Sigma_e(A, B, C, D) := \{w : \mathbb{R} \to \mathcal{W} \mid \exists x \text{ absolutely continuous and } u \in \mathcal{L}^{\text{loc}} \text{ such that } \dot{x}(t) = Ax(t) + Bu(t) \text{ a.e. and } w(t) = Cx(t) + Du(t)\},$$

$$\Sigma'_e(A, B, C, D) := \{w : \mathbb{R} \to \mathcal{W} \mid w \in \mathcal{L}^{\text{loc}} \text{ and there are distributions } x \text{ and } u \text{ such that } \dot{x} = Ax + Bu,\ w = Cx + Du\}.$$

Note that it is actually most natural to consider $\Sigma'_e(A, B, C, D)$ [and not $\Sigma_e(A, B, C, D)$] as the external signals implied by the equations $\dot{x} = Ax + Bu$, $w = Cx + Du$.

The first of the above questions is easily settled on the basis of general principles. In order to do this we need to introduce a familiar concept from the geometric theory of linear systems. The *supremal output-nulling subspace*, $\mathcal{V}^*$, is defined as

$$\mathcal{V}^* := \{x_0 \in \mathcal{X} \mid \exists u \in \mathcal{L}^{\text{loc}} \ni \text{ the trajectory } w \text{ generated by } \dot{x} = Ax + Bu,\ w = Cx + Du,\ x(0) = x_0 \text{ satisfies } w = 0\}.$$

The space $\mathcal{V}^*$ is easily computed from (A, B, C, D) (see [7, in particular Example 4.6] for algorithms and many applications of $\mathcal{V}^*$). We have:

THEOREM 4.3. $\Sigma_i(A, B, C, D)$ is a minimal realization of its own external behavior $\Sigma_e(A, B, C, D)$ *iff* $\mathcal{V}^* = 0$.

Proof (*outline*). $\Rightarrow$: Assume $\mathcal{V}^* \neq 0$; then we can obtain a reduced realization as follows. Define $\mathcal{X}^* := \mathcal{X} \pmod{\mathcal{V}^*}$ and

$$\Sigma_i^* := \{ (x^*, w) : \mathbb{R} \to \mathcal{X}^* \times \mathcal{W} \mid \exists (x, w) \in \Sigma_i(A, B, C, D) \ni x^*(\cdot) = x(\cdot) \pmod{\mathcal{V}^*}\}$$

$\Leftarrow$: Assume that $\Sigma_i(A, B, C, D)$ is a minimal realization of $\Sigma_e(A, B, C, D)$. By Theorem 4.2 this implies that $\Sigma_i(A, B, C, D)$ is equivalent (with a linear bijection) to the canonical past-induced realization (see [1]) of $\Sigma_e(A, B, C, D)$. Consequently $(x, 0) \in \Sigma_i(A, B, C, D)$ will imply $x = 0$, which translates into $\mathcal{V}^* = 0$. ■

The next lemma gives us a sufficient condition under which $\Sigma_e(A, B, C, D) = \Sigma'_e(A, B, C, D)$:

LEMMA 4.4. $\{\ker D \subset \ker B\} \Rightarrow \{\Sigma_e(A, B, C, D) = \Sigma'_e(A, B, C, D)\}$.

Proof. Note that we can assume without loss of generality that

$$\ker\begin{bmatrix} B \\ D \end{bmatrix} = 0$$

(for otherwise, simply eliminate the u's in $\ker\begin{bmatrix} B \\ D \end{bmatrix}$). Hence we need to show that $\{\ker D = 0\} \Rightarrow \{\Sigma_e(A, B, C, D) = \Sigma'_e(A, B, C, D)\}$. By suitably choosing the basis in $\mathcal{U}$ and $\mathcal{W}$ we obtain the following equations describing our system:

$$\dot{x} = Ax + Bu, \quad w_1 = C_1 x + u, \quad w_2 = C_2 x, \text{ and } w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

Solving for u yields $\dot{x} = \tilde{A}x + Bw_1$, $w_2 = C_2 x$, with $\tilde{A} = A - BC_1$. Hence $w \in \mathcal{L}^{\text{loc}}$ implies x absolutely continuous. This yields the result. ■

We will now give the conditions for $\Sigma_i(A, B, C, D)$ to be a minimal realization of $\Sigma'_e(A, B, C, D)$. In order to do this, we introduce another concept from the geometric theory of linear systems. The *supremal $\mathcal{L}_1$-almost output-nulling subspace*, $\mathcal{V}_b^*$, is defined as

$$\mathcal{V}_b^* := \{x_0 \in \mathcal{X} \mid \forall \varepsilon > 0 \ \exists u \in \mathcal{L}^{\text{loc}} \text{ such that the trajectory } w \text{ generated by } \dot{x} = Ax + Bu,\ w = Cx + Du,\ x(0) = x_0, \text{ satisfies } \textstyle\int_{-\infty}^{+\infty} \|w(t)\|\,dt \leqslant \varepsilon\}$$

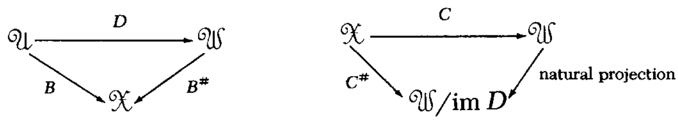
This subspace is a slight generalization of the almost invariant subspaces studied in detail in [8], where algorithms for computing $\mathcal{V}_b^*$ are given. A result which is easily derived from [8] is

$$\mathcal{V}_b^* = \mathcal{V}_{\mathfrak{D}}^* := \{x_0 \in \mathcal{X} \mid \text{there is a distribution } u \text{ with support on } [0, \infty) \text{ such that the distribution } w \text{ generated by } \dot{x} = Ax + Bu,\ w = Cx + Du,\ x(0) = x_0 \text{ satisfies } w = 0\}.$$

By definition the distribution w in the above definition has support on $[0, \infty)$

and is defined as $f + Du + G * u$ with $f: t \in \mathbb{R}^+ \mapsto Ce^{At}x_0$, $*$ convolution, and $G: t \in \mathbb{R}^+ \mapsto Ce^{At}B$. The space $\mathcal{V}^*_{\mathfrak{D}}$ is also studied in [9], where its relevance in many control and linear-systems problems is demonstrated.

The following theorem answers the second question posed earlier on. First, however, some more notation. Assume $\ker D \subset \ker B$. The $B^{\#}$ and $C^{\#}$ are well defined by the commutative diagram



In terms of these, we obtain the following

THEOREM 4.5. *The following conditions are equivalent*:

(i) $\Sigma_i(A, B, C, D)$ *is a minimal realization of* $\Sigma'_e(A, B, C, D)$,
(ii) $\mathcal{V}^*_b = 0$,
(iii) $\ker D \subset \ker B$ *and* $\mathcal{V}^* = 0$,
(iv) $\ker D \subset \ker B$ *and* $(A - B^{\#}C, C^{\#})$ *is observable.*

(It is easy to see that assumption (iv) is independent of $B^{\#}$ provided, of course, $B = B^{\#}D$.)

Proof. (iii) $\Rightarrow$ (i) Follows from Lemma 4.4 and Theorem 4.3.

(iii) $\Leftrightarrow$ (iv): Using the basis in the proof of Lemma 4.4 leads us to consider $\dot{x} = \tilde{A}x + Bw_1$, $w_2 = C_2 x$ with

$$B^{\#}: \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \mapsto Bw_1, \qquad \tilde{A} = A - B^{\#}C, \quad \text{and} \quad C_2 = C^{\#}.$$

The conclusion is an immediate consequence of the definitions of $\mathcal{V}^*$ and of observability.

(i) $\Rightarrow$ (iii): (i) $\Rightarrow$ $\{\mathcal{V}^* = 0\}$ follows from Theorem 4.3. Now choose the basis in $\mathcal{U}$ and $\mathcal{W}$ such that the equations become $\dot{x} = Ax + B_1u_1 + B_2u_2$, $w_1 = C_1x + u_1$, $w_2 = C_2x$. Equivalently, $\dot{x} = \tilde{A}x + B_1w_1 + B_2u_2$, $w_2 = C_2x$ with $\tilde{A} := A - B_1C_1$. Now (i) implies that $\Sigma_e(A, B, C, D) = \Sigma'_e(A, B, C, D)$; hence (looking at $w_1 = 0$) $\Sigma_e(\tilde{A}, B_2, C_2, 0) = \Sigma'_e(\tilde{A}, B_2, C_2, 0)$. By considering the fact that elements in $\Sigma_e(\tilde{A}, B_2, C_2, 0)$ are necessarily absolutely continuous, we see that this equality can only hold if $C_2(Is - A)^{-1}B_2 = 0$. Consequently

$$\langle \ker C_2 | \tilde{A} \rangle := \bigcap_{k=1}^{n} (\tilde{A})^{-k+1} \ker C_2$$

$$\subset \sum_{k=1}^{n} \tilde{A}^{k-1} \operatorname{im} B =: \langle \tilde{A} | \operatorname{im} B_2 \rangle.$$

Hence $\mathcal{V}^* \supset \langle \tilde{A} | \operatorname{im} B_2 \rangle$. Since $\mathcal{V}^* = 0$, this implies $B_2 = 0$, as desired.

(ii) $\Leftrightarrow$ (iii) is worth stating separately:

LEMMA 4.6. $\{\mathcal{V}_b^* = 0\} \Leftrightarrow \{\ker D \subset \ker B$ *and* $\mathcal{V}^* = 0\}$.

Proof. $\Rightarrow$: $\mathcal{V}_b^* = 0$ trivially implies $\mathcal{V}^* = 0$. Assume again

$$\ker \begin{bmatrix} B \\ D \end{bmatrix} = 0.$$

The basis used above gives us $\dot{x} = Ax + B_1 u_1 + B_2 u_2$, $w_1 = C_1 x + u_1$, $w_2 = C_2 x$ for the system equations. From the results in [8] it follows immediately that $\operatorname{im} B_2 \subset \mathcal{V}_b^* = 0$. Hence $\ker D = 0$.

$\Leftarrow$: $w = 0$ in the sense of distributions implies $w = 0$ in the $\mathfrak{L}^{\mathrm{loc}}$ sense. Now $\ker D \subset \ker B$ implies by Lemma 4.4 that the corresponding x will be absolutely continuous. This yields $\mathcal{V}_b^* = \mathcal{V}^*$.

This ends the proof of Lemma 4.6 and Theorem 4.5. ∎

The conditions $\mathcal{V}^* = 0$ and $\mathcal{V}_b^* = 0$ which feature in Theorems 4.3 and 4.5 are reminiscent of strict observability. Indeed, $\mathcal{V}^* = 0$ iff knowledge of the output (w in our case) on $t \geqslant 0$ allows to reconstruct x on $t \geqslant 0$ without knowing the driving input $u \in \mathfrak{L}^{\mathrm{loc}}$. In the literature this is called *strict observability*. The condition $\mathcal{V}_b^* = 0$ allows a similar interpretation. In fact, $\mathcal{V}_b^* = 0$ iff knowledge of the output as a distribution on $t \geqslant 0$ with an unknown input distribution on $t \geqslant 0$ allows one to reconstruct x as a distribution on $t \geqslant 0$. One could call this *distributional strict observability.*

The following corollary follows immediately from conditions (i) and (iv) of Theorem 4.5. Note that in this corollary we consider not just the output y, but the input *and* the output as the external variables, i.e., $w = (u, y)$.

COROLLARY 4.7. *$\dot{x} = Ax + Bu$, $y = Cx + Du$, $w = (u, y)$ is a minimal realization of its own external behavior iff (A, C) is observable.*

It is important to note that in Corollary 4.7 minimality requires only the observability of (A, C) and *not* the reachability of (A, B), contrary to the familiar situation, {minimality} $\Leftrightarrow$ {(A, B) reachable and (A, C) observable}, in the input-output setting with the signals all having half-line support. In our framework, $\dot{x} = Ax$, $w = Cx$ with (A, C) observable is a perfectly well-defined minimal state-space representation of its own external behavior. We shall return to such autonomous systems in Section 8.1.

Given an arbitrary system $\Sigma_i(A, B, C, D)$, which need not be a minimal realization of $\Sigma_e'(A, B, C, D)$, we can reduce it by means of the following algorithm, which is only a slight variation of *Silverman's structure algorithm* [9].

Step 1. Choose the basis in $\mathcal{U}$ and a nonsingular matrix T such that

$$TD = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}.$$

Write conformably

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \qquad T\mathcal{W} = \mathcal{W}_1 \oplus \mathcal{W}_2, \quad \text{and} \quad Tw = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix},$$

to obtain $\dot{x} = Ax + B_1 u_1 + B_2 u_2$, $w_1 = C_1 x + u_1$, $w_2 = C_2 x$. If $\operatorname{im} B_2 = 0$, go to step 4; otherwise go to step 2.

Step 2. Write $\mathcal{X} = \mathcal{X}_1 \oplus \operatorname{im} B_2$ and x conformably as $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, to obtain $\dot{x}_1 = A_{11} x_1 + B_{11} u_1 + B_{12} x_2$, $w_1 = C_{11} x_1 + u_1 + C_{12} x_2$, $w_2 = C_{21} x_1 + C_{22} x_2$.

Step 3. Define

$$A := A_{11}, \quad B := \begin{bmatrix} B_{11} & B_{12} \end{bmatrix}, \quad C := \begin{bmatrix} C_{11} \\ C_{21} \end{bmatrix}, \quad D := T^{-1} \begin{bmatrix} I & C_{12} \\ C_{22} & 0 \end{bmatrix},$$

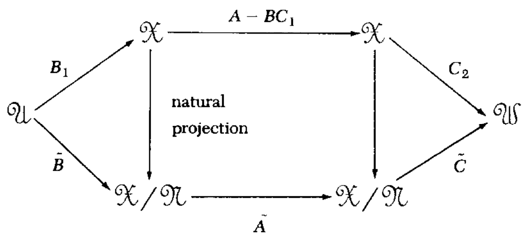
and go to Step 1.

The above loop ends with a system of the form

$$\dot{x} = Ax + B_1 u_1, \quad w_1 = C_1 x + u_1, \quad w_2 = C_2 x_1, \quad Tw = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

Step 4. Compute $\mathcal{N} = \langle \ker C_2 | A - BC_1 \rangle := \bigcap_{k=1}^{n} (A - BC_1)^{-k+1} \ker C_2$, the unobservable subspace of $\dot{x} = (A - BC_1)x$, $w_2 = C_2 x$, and compute

$\tilde{A}$, $\tilde{B}$, $\tilde{C}$ as defined by the commutative diagram



Then

$$\dot{x} = \tilde{A}x + \tilde{B}w_1, \qquad w_2 = \tilde{C}x, \qquad w = T^{-1}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

is a minimal state-space representation of $\Sigma'_e(A, B, C, D)$ (and in fact a minimal i/s/o representation in the sense of Definition 6.1).

5. STATE-SPACE REALIZATIONS OF $\Sigma(R)$

In this section we consider the state-space realization of a system $\Sigma(R)$ as defined in Section 2. We shall establish that there always exists (A, B, C, D) such that $\Sigma(R) = \Sigma_e(A, B, C, D) = \Sigma'_e(A, B, C, D)$ in a minimal way.

THEOREM 5.1. *Let $R \in \mathbb{R}^{r \times q}[s]$ be given. Then there exists (A, B, C, D) such that $\Sigma_i(A, B, C, D)$ is a minimal realization of $\Sigma(R)$ and, in fact, $\Sigma_e(A, B, C, D) = \Sigma'_e(A, B, C, D) = \Sigma(R)$.*

Proof. The proof of this theorem starts from Theorem 3.4 and applies the familiar observer canonical form [6, p. 414]. However, in our case, because of our definition of a system, we need to prove a bit more than mere equality of transfer functions. By Theorem 3.4 there exists a nonsingular matrix T such that $w \in \Sigma(R)$ iff

$$Tw = \begin{bmatrix} u \\ y \end{bmatrix}$$

is governed by $P(d/dt)y = Q(d/dt)u$ with $P(s) = \operatorname{diag}(s^{n_1}, \ldots, s^{n_p}) - S(s)\hat{A}$ and $Q(s) = S(s)\hat{B}$, where $S(s) = \operatorname{diag}(d_1(s), \ldots, d_p(s))$ and $d_i(s) = [1 \;\; s \;\; \cdots$

s^{n_i-1}]. Now define

$$A=\left[\begin{array}{ccc|c|ccc|c|c|ccc|c}
0 & & & & & & & & & & & & \\
1 & & & & & & & & & & & & \\
 & \ddots & & & & & & & & & & & \\
 & & 1 & & & & & & & & & & \\
 & & & & 0 & & & & & & & & \\
 & & & a_1 & 1 & & & a_2 & \cdots & & & & a_p \\
 & & & & & \ddots & & & & & & & \\
 & & & & & & 1 & & & & & & \\
 & & & & & & & & & 0 & & & \\
 & & & & & & & & & 1 & & & \\
 & & & & & & & & & & \ddots & & \\
 & & & & & & & & & & & 1 &
\end{array}\right],$$

n_1-th column $\qquad$ (n_1+n_2)-th column $\qquad$ $(n_1+\cdots+n_p)$-th column

$$\hat{C}=\left[\begin{array}{ccc|c|ccc|c|c|ccc|c}
0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 & \cdots & 0 & 0 \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & 1
\end{array}\right],$$

$$\hat{A}=\begin{bmatrix} a_1 & a_2 & \cdots & a_p \end{bmatrix},$$

$$B=\hat{B},$$

$$C=T^{-1}\begin{bmatrix} 0 \\ \hat{C} \end{bmatrix},$$

$$D=T^{-1}\begin{bmatrix} I \\ 0 \end{bmatrix}.$$

We claim that $\Sigma(R)=\Sigma_e(A,B,C,D)=\Sigma_e'(A,B,C,D)$. This last equality follows immediately from Lemma 4.4. That $\Sigma(R)=\Sigma_e(A,B,C,D)$ follows by direct verification. The idea is as follows. The $(2,\dots,n_1,n_1+2,\dots,n_1+n_2,\dots)$th rows of $\dot{x}=Ax+Bu$, $y=\hat{C}x$ allow us to express x as a linear combination of u, y, and their derivatives up to order $\max_i(n_i-1)$. The $(1,n_1+1,\dots)$th rows of $\dot{x}=Ax+Bu$ become, after substitution of x, exactly the equations $P(d/dt)y=Q(d/dt)u$. From there it is easy to prove that

$$\left(x, T^{-1}\begin{bmatrix} u \\ y \end{bmatrix}\right)\in\Sigma_i(A,B,C,D)$$

implies $P(d/dt)u = Q(d/dt)y$, and conversely, if this equation is satisfied, then with x as constructed above we will obtain

$$\left(x, T^{-1}\begin{bmatrix} u \\ y \end{bmatrix}\right) \in \Sigma_i(A, B, C, D).$$

The minimality of $\Sigma_i(A, B, C, D)$ follows immediately from Theorem 4.5(iv). ■

The combination of Theorems 3.5 and 5.1 yields an algorithm for obtaining a minimal realization of $\Sigma(R)$. This requires making R row-proper by premultiplication by a unimodular matrix and deriving a standard observable realization from there. We make no claims as to the efficiency of this procedure. What we would actually like to develop is a generalization, starting directly from R, of Fuhrmann's elegant "polynomial model" realization [3].

It is useful to note that realizing $\Sigma(R)$ simply requires finding (A, B, C, D) such that $[C\ D]\ker[Is - A \mid -B] = \ker R(s)$ [11]. This condition for $s \in \mathbb{C}$ guarantees only that $\Sigma_e'(A, B, C, D) = \Sigma(R)$. If we add the condition at $s = \infty$, i.e., if we demand that $[C\ D]\lim_{s \to \infty}\ker[Is - A \mid -B] = \operatorname{im} D = \operatorname{im}_{s \to \infty} \ker R(s)$ (convergence to be understood in the Grassmann sense), then we may conclude that $\Sigma_e(A, B, C, D) = \Sigma(R)$.

6. INPUT-STATE-OUTPUT REPRESENTATIONS

The analogue of Definition 3.4 for state-space systems is the following:

DEFINITION 6.1. A system defined by

$$\dot{x} = \tilde{A}x + \tilde{B}u, \qquad y = \tilde{C}x + \tilde{D}u, \qquad w = T^{-1}\begin{bmatrix} u \\ y \end{bmatrix}$$

with T a nonsingular matrix is said to be in *input-state-output* (i/s/o) *form* with u in the *input*, x the *state*, and y the *output*. We will denote such a system by $\{(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}), T\}$. Note that for an i/s/o system we have, by Lemma 4.4, $\Sigma_e'(\cdot) = \Sigma_e(\cdot)$, and hence we need not worry about smoothness.

From Theorem 4.5, it follows that an i/s/o system is minimal iff $(\tilde{A}, \tilde{C})$ is observable. As an immediate consequence of the proof of Theorem 5.1 and

Proposition 3.3 we have:

THEOREM 6.1. *Every system $\Sigma(R)$ admits a minimal i/s/o representation (i.e., given any R, there exist $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ with $(\tilde{A}, \tilde{C})$ observable, and a nonsingular T such that the system defined in Definition 6.1 generates exactly the w's in $\Sigma(R)$). The same holds consequently for $\Sigma(R_1, R_2)$ and $\Sigma_e'(A, B, C, D)$.*

As may be expected, there is an intimate relation between i/o and i/s/o representations of Σ(R). Indeed, we have

THEOREM 6.2. *Let $\Sigma(R)$ be given and T be a nonsingular matrix. Then $\{\exists(P, Q)$ such that $\{(P(s), Q(s)), T\}$ is an i/o representation of $\Sigma(R)\} \Leftrightarrow \{\exists(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ such that $\{(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}), T\}$ is an i/s/o representation of $\Sigma(R)\}$.*

Proof. $\Rightarrow$ follows directly from the proof of Theorem 5.1

$\Leftarrow$: It suffices to show that the external behavior (u, y) of $\dot{x} = \tilde{A}x + \tilde{B}u$, $y = \tilde{C}x + \tilde{D}u$ equals that of $P(d/dt)y = Q(d/dt)u$ for suitable P, Q's as in Example 3.1. We do not show this in detail. The proof proceeds as follows: first note that without loss of generality we can assume $(\tilde{A}, \tilde{C})$ observable (otherwise reduce x to x (mod $\mathfrak{N}$) with $\mathfrak{N} = \langle \ker \tilde{C} | \tilde{A} \rangle := \cap_{k=1}^{n} \tilde{A}^{-k+1} \ker \tilde{C}$, the unobservable subspace of $(\tilde{A}, \tilde{C})$. Then use a basis transformation in the state space to put $(\tilde{A}, \tilde{C})$ into the observer canonical form as used in the proof of Theorem 5.1. This then defines, as shown in that proof, the equivalent equations $P(d/dt)y = Q(d/dt)u$. ∎

As an immediate corollary of Theorems 6.2 and 3.5, we obtain

COROLLARY 6.3. $\Sigma(R)$ *admits a minimal (i.e., observable) i/s/o representation with T a permutation matrix, or alternatively with $\tilde{D} = 0$.*

REMARKS. Theorem 3.5 and the above theorem allow us to draw a number of interesting conclusions:

REMARK 6.4. Any system $\Sigma(R)$ [and hence $\Sigma(R_1, R_2)$ and $\Sigma_e'(A, B, C, D)$] admits an i/o and an i/s/o representation. Consequently for this class of systems the causality issue does not arise. It is merely a representation *fact*.

REMARK 6.5. Since T may be chosen as a permutation matrix, we may conclude that we can always partition the vector w *componentwise* into inputs and outputs. Thus an i/o representation is a matter of partitioning the external variables correctly.

REMARK 6.6. If we do not insist on a componentwise partitioning of w, then we can even assume strict properness of the transfer function. Thus strict causality is also a representation result.

REMARK 6.7. We should expect global representation results such as those obtained in Theorems 3.5 and 6.2 to be limited to linear and time-invariant systems. In fact it can be argued [10, 1] that one should not expect i/s/o representations in a nonlinear differential-geometric context. The starting point $\dot{x} = f(x, u)$, $y = g(x, u)$ of much of control theory is more restricted than is often realized.

REMARK 6.8. Assume that $\Sigma(R)$ is given, and consider now the following subspace of $\mathscr{W}$:

$$\mathscr{W}_u := \Big\{ w_0 \in \mathscr{W} \,|\, \exists w \in \Sigma(R) \text{ with } w(t) = 0 \text{ for } t < 0,$$

$$w \text{ continuous for } t \geqslant 0, \text{ and } \lim_{t \to 0^+} w(t) = w_0 \Big\}.$$

Then clearly if $\{(P(s), Q(s)), T\}$ or $\{(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}), T\}$ is an i/o or an i/s/o representation of $\Sigma(R)$, we have

$$T\mathscr{W}_u = \operatorname{im}\begin{bmatrix} I \\ \lim_{s \to \infty} P^{-1}(s)Q(s) \end{bmatrix} = \operatorname{im}\begin{bmatrix} I \\ \tilde{D} \end{bmatrix}.$$

From this it immediately follows that $\dim \mathscr{U} = \dim \mathscr{W}_u$. Hence $\dim \mathscr{U}$ and $\dim \mathscr{Y}$ are invariants of $\Sigma(R)$: the *number* of input and output variables is intrinsically defined, while the variables themselves are not. However, if we concentrate on strictly causal representations ($\tilde{D} = 0$), then, since $\mathscr{U} = T\mathscr{W}_u$, we obtain also that the input space is intrinsically defined, while for the output space $\mathscr{Y}$ we can take $\mathscr{Y} = T\mathscr{W}_y$, with $\mathscr{W}_y$ any complement of $\mathscr{W}_u$.

7. THE TRANSFORMATION GROUP ASSOCIATED WITH A MINIMAL REALIZATION

The following theorem gives all minimal realizations having the same external behavior:

THEOREM 7.1. *Assume that $\Sigma_i(A, B, C, D)$ is a minimal realization of its own external behavior $\Sigma_e(A, B, C, D)$, with*

$$\ker\begin{bmatrix} B \\ D \end{bmatrix} = 0.$$

Then all such minimal realizations are obtained by the transformation group

$$(A, B, C, D) \underset{\substack{\det S \neq 0 \\ \det R \neq 0}}{\overset{F,S,R}{\mapsto}} (S(A+BF)S^{-1}, SBR, (C+DF)S^{-1}, DR).$$

Proof. By Theorem 4.2 all minimal realizations are linearly equivalent. Let S be a nonsingular matrix. We would hence like to find (A', B', C', D') such that $(Sx, w) \in \Sigma_i(A', B', C', D') \Leftrightarrow (x, w) \in \Sigma_i(A, B, C, D)$. Now,

$$(x, w) \in \Sigma_i(A, B, C, D) \quad \Leftrightarrow \quad \begin{bmatrix} \dot{x} - Ax \\ w - Cx \end{bmatrix} \in \operatorname{im}\begin{bmatrix} B \\ D \end{bmatrix}$$

$$\Leftrightarrow \quad \begin{bmatrix} S\dot{x} - SAS^{-1}Sx \\ w - CS^{-1}Sx \end{bmatrix} \in \operatorname{im}\begin{bmatrix} SB \\ D \end{bmatrix}.$$

Hence

$$\begin{bmatrix} S\dot{x} - A'Sx \\ w - C'Sx \end{bmatrix} \in \operatorname{im}\begin{bmatrix} B' \\ D' \end{bmatrix} \quad \Leftrightarrow \quad \begin{bmatrix} S\dot{x} - SAS^{-1}Sx \\ w - CS^{-1}Sx \end{bmatrix} \in \operatorname{im}\begin{bmatrix} SB \\ D \end{bmatrix}$$

In other words, defining

$$\begin{bmatrix} S\dot{x} \\ w \end{bmatrix} =: r, \qquad Sx =: z,$$

$$\operatorname{im}\begin{bmatrix} B' \\ D' \end{bmatrix} =: \mathfrak{L}', \qquad \operatorname{im}\begin{bmatrix} SB \\ D \end{bmatrix} =: \mathfrak{L},$$

$$\begin{bmatrix} A' \\ C' \end{bmatrix} = M', \qquad \begin{bmatrix} SAS^{-1} \\ CS^{-1} \end{bmatrix} = M,$$

we see that $(r - M'z) \in \mathfrak{L}' \Leftrightarrow (r - Mz) \in \mathfrak{L}$. From this it follows that $\mathfrak{L}' = \mathfrak{L}$ and that $Mz = M'z \pmod{\mathfrak{L}}$. Now, $M = M' \pmod{\mathfrak{L}}$ iff there exists F such that

$$M' = M + \begin{bmatrix} SB \\ D \end{bmatrix} F$$

and, since $\begin{bmatrix} SB \\ D \end{bmatrix}$ is injective, $\mathfrak{L}' = \mathfrak{L}$ iff there exists R, $\det R \neq 0$, such that

$$\begin{bmatrix} SB \\ D \end{bmatrix} R = \begin{bmatrix} B' \\ D' \end{bmatrix}. \qquad \blacksquare$$

The above theorem shows that in our setup the relevant transformation group contains not only the state-space isomorphism group, but also the feedback group. This is basically due to the fact that feedback does not change the set of possible trajectories produced by a system. In particular it implies that if (A, B) is reachable, then we may always obtain a minimal realization with (A, B) in Brunovsky canonical form [7, p. 118].

As we have seen, we can always choose an i/s/o representation such that $\tilde{D} = 0$. From Remark 6.8 it follows that this corresponds to writing $T\mathfrak{W}$ as $\mathfrak{U} \oplus \mathfrak{Y}$ with $\mathfrak{U} = T\mathfrak{W}_u$. Considering only these representations yields the following corollary.

Corollary 7.2. *Let $\{(\tilde{A}, \tilde{B}, \tilde{C}, 0), T\}$ be a minimal (i.e., observable) i/s/o representation of its external behavior. Then all such realizations are obtainable by the action of the transformation group*

$$(\tilde{A}, \tilde{B}, \tilde{C}) \underset{\substack{\det S \neq 0 \\ \det R_1 \neq 0 \\ \det R_2 \neq 0}}{\overset{F, S, R_1, R_2}{\mapsto}} \left(S(\tilde{A} + \tilde{B}F\tilde{C})S^{-1}, S\tilde{B}R_1, R_2\tilde{C}S^{-1}\right),$$

$$T \mapsto T \begin{bmatrix} R_1 & FR_2^{-1} \\ 0 & R_2^{-1} \end{bmatrix}.$$

8. COMMENTS

8.1. Autonomous and Reachable Systems

In the context of our definitions it is possible to view the fact that a system is autonomous or reachable as *external* properties. The system $\Sigma(R)$ is said to be *autonomous* if $\exists f: \mathcal{W}^{(-\infty,0)} \to \mathcal{W}^{[0,\infty)}$ such that $w \in \Sigma(R) \Leftrightarrow w^+ = f(w^-)$. It is said to be *reachable* (from 0) if for all $w \in \Sigma(R)$ there exists $\tilde{w} \in \Sigma(R)$ with $w^+ = \tilde{w}^+$ and $\tilde{w}^-$ of compact support, i.e., the family of future trajectories equals those which follow trajectories which are zero in the far past. In an autonomous system we have $\mathcal{W}_u = 0$. In terms of an i/o representation, an autonomous system has $Q = 0$, while a reachable one has P and Q left coprime. An autonomous system admits a realization of the form $\dot{x} = Ax$, $w = Cx$ [with (A, C) observable $\Leftrightarrow$ minimality], while $\Sigma(R)$ is reachable iff it has a realization $\Sigma_i(A, B, C, D)$ with (A, B) a reachable pair, in which case all its minimal realizations or i/s/o representations will be reachable. We can always define the *reachable component* of $\Sigma(R)$. The reachable component Σr is defined by taking a minimal realization $\Sigma_i(A, B, C, D)$ of $\Sigma(R)$ and considering the external behavior of

$$\Sigma_i^r(A, B, C, D) := \{ (x, w) \in \Sigma_i(A, B, C, D) | x(t) \in \langle A | \operatorname{im} B \rangle \ \forall t \},$$

which is still a state-space system. By Proposition 3.3, $\Sigma_r = \Sigma(R_r)$ for some R_r. Finally $\Sigma(R)$ can always be written as $\Sigma(R) = \Sigma_r + \Sigma_a$ with Σ_a autonomous.

8.2. *The Algebraic-Geometric Structure*

It is trivial to see that all what we have said up to now also holds for $\mathcal{W} = \mathbb{C}^q$. In this section we assume that we are working over $\mathbb{C}$. A system $\Sigma(R)$ defines a map S from $\mathbb{C}$ into $\mathcal{G}^q$, the set of subspaces of $\mathbb{C}^q$, defined by $S(s) := \ker R(s)$. The set $\mathcal{B} := \{(s, w) | s \in \mathbb{C},\ w \in S(s)\}$ obviously defines a fibration over the base space $\mathbb{C}$. However, since $\dim S(s)$ need not be constant, $\mathcal{B}$ is in general not a vector bundle. Nevertheless, since $S(s)$ is the kernel of a matrix polynomial, $\mathcal{B}$ has some nice mathematical structure. It is what in algebraic geometry is called an *algebraic coherent sheaf*. Now, if $S(s)$ has constant dimension, there exists a bijection from $\mathcal{B}$ to $\Sigma(R)$, and moreover $\mathcal{B}$ is actually a vector bundle over $\mathbb{C}$. Furthermore, $\dim S(s)$ is constant for all but a finite set of points s. Let us denote by m this "normal" dimension of $S(s)$, and by $\mathbb{C}_a$ the points where $\dim S(s) > m$.

This algebraic-geometric structure is exploited in [11] in order to obtain elegant interpretations of a number of system-theoretic facts, in the spirit of [12]. For example, $\{\dim S(s) = m\ \forall s \in \mathbb{C}\} \Leftrightarrow \{\Sigma(R)$ is reachable$\}$, and the points in $\mathbb{C}_a$ correspond to the unreachable modes of any minimal i/s/o realization of $\Sigma(R)$; m equals the number of inputs in any i/o or i/s/o representation of $\Sigma(R)$. Now, $\lim_{s \to \infty} S(s)$ is well defined in $\mathcal{G}_m^q$, the Grassmannian of m planes in $\mathbb{C}^q$, and equals $\mathcal{W}_u$, the intrinsic input space.

Replacing $S(s')$ by $\lim_{s \to s'} S(s)$ at points $s' \in \mathbb{C}_a$ corresponds to replacing $\Sigma(R)$ by its reachable component Σ_r. Let $R_r(s)$ be such that $\Sigma_r = \Sigma(R_r)$. Assume now that $\Sigma(R)$ is reachable (or that R is replaced by R_r), and by defining $S(\infty) := \mathcal{W}_u$, extend $\mathcal{B}$ to an algebraic vector bundle $\mathcal{B}'$ over $\mathbb{P}$, the projective line. The problem of realization—i.e., finding (A, B, C, D) such that $[C\ D]\ker[Is - A \mid -B] = S(s)$—corresponds to unfolding $\mathcal{B}'$. By a theorem of Grothendieck, every algebraic vector bundle over $\mathbb{P}$ is isomorphic to a direct sum of line bundles. The *Chern numbers* of these line bundles are precisely the input Kronecker indices of any minimal i/s/o representation of $\Sigma(R)$. The Chern numbers of $\mathcal{B}'' := \{(s, w) | s \in \mathbb{P},\ w \in (\ker R(s))^{\perp}\}$ are the output Kronecker indices of any minimal i/s/o representation of $\Sigma(R)$ with $\tilde{D} = 0$. Finally, all this induces a *bijection* from the set of algebraic (and hence holomorphic) vector bundles over the Riemann sphere with positive Chern numbers to the reachable linear time-invariant finite-dimensional systems. For details and proofs of all this, see [11].

8.3. *Smoothness*

We have chosen to interpret $\Sigma(R)$ and $\Sigma(R_1, R_2)$ with $w \in \mathfrak{L}^{\mathrm{loc}}$, and with the auxiliary variables, the derivatives, and the equations to be interpreted in the sense of distributions. It was only when considering $\Sigma_e(A, B, C, D)$ that x was assumed to be absolutely continuous and $u \in \mathfrak{L}^{\mathrm{loc}}$. If, for some reason, we require more smoothness (e.g. that $w^{(k)} \in \mathfrak{L}^{\mathrm{loc}}$ or $\xi^{(l)} \in \mathfrak{L}^{\mathrm{loc}}$), then we should simply add the equations $w^{(k)} = w'$, $\xi^{(l)} = \xi'$ and apply the results with $\begin{bmatrix} w \\ w' \end{bmatrix}$ as external variables and $\begin{bmatrix} \xi \\ \xi' \end{bmatrix}$ as auxiliary variables. However in this case we shall not always obtain an i/o or an i/s/o representation of the original w-system. In fact $\Sigma_e(A, B, C, D)$, for example, need not have an i/s/o representations.

It is interesting to note that a theory which assumes all signals to be C^{∞} would yield essentially identical representation results (like Theorems 4.5, 5.1, and 6.2) to those obtained from the distributional viewpoint taken here. The only difference is that our axiom of state $\dot{x} = Ax + Bu$, $w = Cx + Du$ would then not be valid any more. However, this provides additional evidence for the point of view which takes $\Sigma(R) = \Sigma_e'(A, B, C, D)$ as the basic equality to demand in realization theory. By Theorem 4.5 this then supports the important role which $\mathcal{V}_b^*$ and almost controlled invariant subspaces can be expected to play.

8.4. *An Example*

The following (trivial) example serves to illustrate some of the ideas of the paper. Take $\mathcal{W} = \mathbb{R}^2$, let $p_1, p_2 \in \mathbb{R}[s]$, not both zero, and consider the system, with

$$w=\begin{bmatrix} w_1 \\ w_2 \end{bmatrix},$$

described by

$$p_1\left(\frac{d}{dt}\right)w_1=p_2\left(\frac{d}{dt}\right)w_2.$$

We regard this as defining a family of trajectories $(w_1(\cdot), w_2(\cdot))$: simply those which satisfy this differential equation in the sense explained earlier. What should we choose as the input? This is very easy to decide for the system at hand. Let $n_1:=\text{degree } p_1$ and $n_2:=\text{degree } p_2$. If we consider componentwise partitions of w into inputs and outputs, then, if $n_1>n_2$, w_1 is the output and w_2 the input. Conversely, if $n_1<n_2$, w_2 is the output and w_1 is input. If $n_1=n_2$, then w_1 or w_2 may be chosen as input, and the other becomes the output. In this case we will have a not strictly proper transfer function, however. If we want a strictly proper transfer function, then we should choose u and y as linear combinations of w_1 and w_2. In the case at hand, if $p_1(s)=r_1s^n+\cdots$ and $p_2(s)=r_2s^n+\cdots$, with $r_1r_2\neq 0$, then it is easily verified that if we set $y=\alpha(r_1w_1-r_2w_2)$, $\alpha\neq 0$, and $u=r_1'w_1-r_2'w_2$ with $r_1r_2\neq r_1'r_2'$, then u and y are related by an equation of the type

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

with $p,q\in\mathbb{R}[s]$, p of degree n, and q of smaller degree. This is an i/o representation of the system with

$$\begin{bmatrix} r_1' & -r_2' \\ \alpha r_1 & -\alpha r_1 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}=\begin{bmatrix} u \\ y \end{bmatrix}$$

and u, y related by the above differential equation.

Finally, a few remarks on the state-space realization of this system. Its minimal linear time-invariant realizations will *always* have state-space dimension exactly $n:=\max(n_1,n_2)$. Common factors will have as a consequence that a minimal realization is not reachable. In fact, the number of common factors of p_1 and p_2 equals precisely the difference between the dimension of the minimal state space and the reachable subspace. It is easy to write down a minimal realization. Assume $n=n_1\geqslant n_2$. Let $p_1(s)=p_ns^n+$

$p_{n-1}s^{n-1} + \cdots + p_0$ and $p_2(s) = q_n s^n + q_{n-1}s^{n-1} + \cdots + q_0$. Let

$$\frac{p_2(s)}{p_1(s)} = b_0 + \frac{b_1}{s} + \frac{b_2}{s^2} + \cdots$$

be the Laurent series expansion of p_2/p_1. Then choosing

$$A = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \frac{-p_0}{p_n} & \frac{-p_1}{p_n} & \cdots & \frac{-p_{n-1}}{p_n} \end{bmatrix}, \qquad B = \begin{bmatrix} b_1 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}, \quad D = b_0$$

yields, with $\dot{x} = Ax + Bw_2$, $w_1 = Cx + Dw_2$, a minimal state-space (in fact i/s/o) realization. In particular, if $p_2 = 0$, we see that the system reduces to the autonomous w_1 governed by $p_1(d/dt)w_1 = 0$ and the completely unrelated free input w_2. Its minimal realization is $\dot{x} = Ax$, $w_1 = Cx$, with (A, C) as above, and w_2 arbitrary (in $\mathfrak{L}^{\text{loc}}$). Clearly the only w_1 trajectory with half-line support is then the null trajectory.

9. CONCLUSIONS

In this paper we have provided a theory for the input-output and state-space modeling of systems described by higher-order differential equations $f(w, \dot{w}, \ldots, w^{(k)}) = 0$, with f linear. The main conclusion is that such systems always admit a familiar i/o and an i/s/o representation. Hence "causality" is a matter of choosing the inputs and outputs appropriately.

All this yields ample additional evidence for the fact that the standard input-output framework of Kalman [13] provides essentially an impeccable setting for linear systems theory. Indeed, even the seemingly more general context $R(d/dt)w = 0$ can be reduced to it in a precise way. In some sense the main new element which this more general vantage point brings out (other than an intellectually perhaps more pleasing framework) is that one should not take reachability for granted. Indeed, in our more general setting it is *not* a consequence of working with a minimal realization. This is contrary to what can be obtained in the input-output setting.

When writing down the dynamical equations modeling a system, we shall usually end up with a set of equations of the type $f_1(w, \dot{w}, \ldots, w^{(k)}, \xi, \dot{\xi}, \ldots, \xi^{(l)}) = 0$, where the w's are the variables which are being modeled and the ξ's are auxiliary variables introduced in order to facilitate writing down equations for the w's. If f_1 is linear, then we can reduce this situation to that of the previous paragraph, i.e., the auxiliary variables may always be eliminated from the equations.

Of course, there may be situations where, for reasons having to do with the background of the problem, one wants (for example) certain components of w to be inputs or outputs, or certain or all ξ's to be causally related to the system inputs, or the relation between certain of the w's or the ξ's to be a (possibly nonproper) transfer function (as for example in Rosenbrock's system matrices). This will in general lead to compatibility conditions between these requirements and the model, i.e., these requirements add structure to the system. The framework presented in the present paper demonstrates however that in principle such issues need not arise in a general theory of finite-dimensional linear time-invariant systems.

REFERENCES

1 J. C. Willems, System theoretic models for the analysis of physical systems, *Ricerche di Automatica* (Special Issue on Systems Theory and Physics) 10(2):71–106 (1979).

2 J. C. Willems, Modeling and representation of dynamical systems defined in terms of external variables, in *Proceedings of the 20th IEEE Conference on Decision and Control*, 1981, pp. 1326–1330.

3 P. A. Fuhrmann, Algebraic system theory: an analyst's point of view, *J. Franklin Inst.* 301:521–540 (1976).

4 H. H. Rosenbrock, *State Space and Multivariable Theory*, Wiley, New York, 1970.

5 W. A. Wolovich, *Linear Multivariable Systems*, Springer, New York, 1974.

6 T. Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, N.J., 1980.

7 W. M. Wonham, *Linear Multivariable Control: A Geometric Approach* 2nd ed. Springer, New York, 1979.

8 J. C. Willems, Almost invariant subspaces: an approach to high gain feedback design—Part 1: Almost controlled invariant subspaces, *IEEE Trans. Automat. Control* AC-26:235–252 (1981).

9 M. L. J. Hautus and L. M. Silverman, System structure and singular control, *Linear Algebra Appl.* 50:369–402 (1983).

10 R. W. Brockett, Control theory and analytical mechanics, in *Geometric Control Theory* (C. Martin and R. Hermann, Eds.), Math. Sci. Press. Brookline, Mass, 1977, pp. 1–46.

11 A. J. van der Schaft, System theoretic descriptions of physical systems, Doctoral Diss., Mathematics Dept., Univ. of Groningen, The Netherlands, 1983.

12 C. Martin and R. Hermann, Applications of algebraic geometry to systems theory: The McMillan degree and Kronecker indices of transfer functions as topological and holomorphic system invariants, *SIAM J. Control and Optim.* 16:743–755 (1978).

13 R. E. Kalman, P. L. Falb, and M. Arbib, *Topics in Mathematical System Theory*, McGraw-Hill, 1969.

Received 25 October 1982