1

# Internal model principles for observers

Jochen Trumpf, Member, IEEE, Harry L. Trentelman, Senior Member, IEEE, and Jan C. Willems, Life Fellow, IEEE

Abstract—This paper deals with the observer problem for dynamical systems in a behavioral context. We are given a dynamical system together with a partition of the system variables into a set of known or measured variables and a set of unknown, to be estimated variables. The observer problem is to find a system that produces an estimate of the unknown variables on the basis of the known or measured variables. For a given plant and partition, we establish a characterization of all error behaviors that can be achieved by interconnecting the plant with some observer. The main result of this paper is a very general, behavioral formulation of an internal model principle for observers. We will show that a *nonintrusive* observer achieves a stable error behavior if and only if, in addition to a detectability condition on the observer, the observer behavior contains the anti-stabilizable part of the plant behavior.

## I. INTRODUCTION

Dynamical systems are mathematical models that describe the evolution in time of a set of variables. Often some of these system variables are known, or accesible for measurement, while others are unknown and to be estimated. Natural questions are then whether these unknown variables can be reconstructed or estimated on the basis of the known or measured variables, and how to produce these reconstructed variables or estimates. This general problem has been studied extensively in the systems and control literature, and is often referred to as the observer problem. A major part of the literature on observer design is concerned with finite-dimensional, linear, time-invariant, input-output systems in state space form. In general the problem here is to reconstruct or estimate a specific (unknown) set of output variables, e.g. a particular linear function of the state, using the values of a different set of additional system variables, like the (known) input trajectories and/or the values of a measured output. This problem dates back to Luenberger [1].

More recently, the observer problem has been studied in the context of the behavioral approach to systems and control. A distinguishing feature of the behavioral approach is that it uses dynamical systems in which the system variables are not explicitly labeled as inputs or outputs. In principle, all variables are treated on an equal footing. Also, the models do not need to be described in state space form. Rather, in the behavioral approach a dynamical system is defined by the whole *set* of system trajectories that are allowed by the laws

J. C. Willems, deceased, was with the Department of Electrical Engineering, Katholieke Universiteit Leuven, B-3001, Leuven-Heverlee, Belgium. of the system. This set of trajectories is called *the behavior* of the system, and is considered to be the core of the dynamical system. In this context, the observer problem becomes how to reconstruct/estimate from a given set of known or observed components of the system variable a complementary set of unknown components of that system variable.

A concise introduction to the observer problem using the behavioral approach has been given in [2]. There, the concept of observer was defined, and conditions were derived for their existence. Also, the results obtained were applied to state estimation for input/state/output systems. In [3], additional results were obtained in the context of discrete-time behaviors, specifically on the existence of deadbeat observers. The authors also demonstrated how the general behavioral results can be specialized to other types of observer problems such as unknown input observers or fault detection and isolation.

The aim of the present paper is to contribute to the further theoretical development of the behavioral approach to systems and control, and in particular to the behavioral theory of observer design. For the sake of clarity we stress that in the behavioral theory of observers there is no requirement that the to be estimated variables are state variables. In fact, an important insight to be gained from approaching the observer problem in this setting is that the derived conditions are *independent* from the type of variable being observed or estimated. As was argued extensively in [4], any reasonable theoretical framework for control systems analysis and design should require the possibility to deal with system variables that are not necessarily labeled as inputs, states or outputs. In such a context, observer problems as studied in this paper emerge in a natural way.

In the present paper, we will introduce the notion of *achievability* in the context of observer design. Given a plant behavior, we will explicitly characterize all error behaviors that can be achieved by interconnecting the plant with some observer. The notion of achievability has been used extensively in the context of control by behavioral interconnection, see [5], [6], where both the terminology 'achievability' and 'implementability' was used. Necessary and sufficient conditions for the existence of tracking, asymptotic, and exact observers (cf. [2]) will follow readily from our characterization, providing an alternative way to derive these conditions.

The main contribution of this paper is a behavioral formulation of a so-called *internal model principle for observers*. It will be shown that a *nonintrusive* observer can only lead to a reasonable error behavior if it contains a relevant part of the plant behavior. More precisely, we will show that a *nonintrusive* observer achieves a stable error behavior if and only if, in addition to a detectability condition on the observer, the observer behavior contains the *anti-stabilizable part* of the

J. Trumpf is with the Research School of Engineering, Australian National University, Canberra ACT 0200, Australia. Email: Jochen.Trumpf@anu.edu.au

H. L. Trentelman is with the Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, 9700 AV Groningen, The Netherlands. Email: h.l.trentelman@math.rug.nl

plant behavior. We also formulate refinements of this internal model principle for tracking observers and for exact observers.

The results are entirely *representation free* and hence could be specialized to any preferred form of (equation) representation. We show how to do this for kernel representations.

We will conclude this paper with two worked out examples, namely an application of the behavioral results to unknown input observer design for descriptor systems, and another application to the case of strictly proper input-output systems represented in state space form. In the latter case we give an interpretation of the internal model principle in terms of a simulation relation between the system and the observer.

To conclude this section, some words on notation and nomenclature used. We use the standard symbols for the fields of real and complex numbers  $\mathbb{R}$  and  $\mathbb{C}$ . We use  $\mathbb{R}^n$ ,  $\mathbb{R}^{n \times m}$ , etc. for the real linear spaces of vectors and matrices with components in  $\mathbb{R}$ . Often, the notation  $\mathbb{R}^w$ ,  $\mathbb{R}^{w_1}$ , ... is used if w,  $w_1$ , ... denote typical elements of that vector space, or typical functions taking their values in that vector space.  $\mathfrak{C}^{\infty}(\mathbb{R}, \mathbb{R}^w)$ will denote the set of infinitely often differentiable functions from  $\mathbb{R}$  to  $\mathbb{R}^w$ .  $\mathbb{R}[\xi]$  denotes the ring of polynomials in the indeterminate  $\xi$  with real coefficients. We use  $\mathbb{R}^{n \times m}[\xi]$  for the space of matrices with components in  $\mathbb{R}[\xi]$ . Elements of  $\mathbb{R}^{n \times m}[\xi]$  are called *real polynomial matrices*.

This paper is structured as follows. Section II recalls some relevant notions from behavioral system theory. Sections III and IV contain results on achievable error behaviors and existence of observers, respectively. The internal model principle is discussed in Section V and Section VI contains an application to unknown input observer design for descriptor systems. Section VII treats the special case of strictly proper input-output systems in state space form. Short conclusions are provided in Section VIII.

A preliminary version of some of the results in this paper was presented at the CDC 2011 [7].

#### **II. PRELIMINARIES**

In the behavioral approach a dynamical system is given by a triple  $\Sigma = (T, W, \mathfrak{B})$ , where T is the time axis, W is the signal space, and the behavior  $\mathfrak{B}$  is a subset of  $W^T$ , the set of all functions from T to W. Since  $\mathfrak{B}$  implicitly carries the information about the choice of T and W, it is common to not carefully distinguish between the system and its behavior. We will allow ourselves to think of the behavior as (defining) a system, and hence often refer to "the system  $\mathfrak{B}$ " in the sequel.

The basic idea of interconnection in this framework is very simple. If  $\Sigma_1 = (T, W, \mathfrak{B}_1)$  and  $\Sigma_2 = (T, W, \mathfrak{B}_2)$  are two dynamical systems with the same time axis and the same signal space, then the *full interconnection*  $\Sigma_1 \wedge \Sigma_2$  of  $\Sigma_1$  and  $\Sigma_2$ is defined as the dynamical system  $(T, W, \mathfrak{B}_1 \cap \mathfrak{B}_2)$ , i.e. the system whose behavior is equal to the set-theoretic intersection of the behaviors  $\mathfrak{B}_1$  and  $\mathfrak{B}_2$ . We speak of full interconnection since the entire variable w of  $\mathfrak{B}_1$  is shared with  $\mathfrak{B}_2$  in the interconnection.

In the present paper, interconnections will in general take place through pre-specified components of the manifest variable. In that case, we speak of *partial interconnection*. Let  $\Sigma_1 = (T, W_1 \times C, \mathfrak{B}_1)$  and  $\Sigma_2 = (T, W_2 \times C, \mathfrak{B}_2)$  be two dynamical systems with the same time axis. We assume that the signal spaces  $W_1 \times C$  and  $W_2 \times C$  of  $\Sigma_1$  and  $\Sigma_2$ , respectively, are product spaces, with the factor C in common. Correspondingly, trajectories of  $\mathfrak{B}_1$  are denoted by  $(w_1, c)$  and trajectories of  $\mathfrak{B}_2$  by  $(w_2, c)$ . We define the *interconnection* of  $\Sigma_1$  and  $\Sigma_2$  through c as the dynamical system

$$\Sigma_1 \wedge_c \Sigma_2 := (T, W_1 \times W_2 \times C, \mathfrak{B}_1 \wedge_c \mathfrak{B}_2)$$

with interconnected behavior

$$\mathfrak{B}_1 \wedge_c \mathfrak{B}_2 := \{ (w_1, w_2, c) \mid (w_1, c) \in \mathfrak{B}_1 \text{ and } (w_2, c) \in \mathfrak{B}_2 \}.$$

The behaviors  $\mathfrak{B}_1$  and  $\mathfrak{B}_2$  in this case only share the variable c, which is called the *interconnection variable*.

In this paper we will restrict ourselves to linear timeinvariant differential systems. A *linear time-invariant differential system* is a dynamical system with time axis  $T = \mathbb{R}$ , and whose signal space W is a finite dimensional Euclidean space, say,  $\mathbb{R}^{\texttt{w}}$ ; correspondingly, the manifest variable is then given as  $w = \operatorname{col}(w_1, w_2, \ldots, w_{\texttt{w}})$ ; the behavior  $\mathfrak{B}$  is a linear subspace of  $\mathfrak{C}^{\infty}(\mathbb{R}, \mathbb{R}^{\texttt{w}})$  consisting of all solutions of a set of higher order, linear, constant-coefficient differential equations, i.e., there exists a positive integer g and a polynomial matrix  $R \in \mathbb{R}^{\texttt{g} \times \texttt{w}}[\xi]$  such that

$$\mathfrak{B} = \{ w \in \mathfrak{C}^{\infty}(\mathbb{R}, \mathbb{R}^{\mathsf{w}}) \mid R(\frac{\mathrm{d}}{\mathrm{d}t})w = 0 \}.$$

The set of linear time-invariant differential systems with manifest variable w taking its value in  $\mathbb{R}^w$  is denoted by  $\mathfrak{L}^w$ . See [8] for a representation-free characterization of this class of systems. In this paper, behaviors are always linear subspaces of  $\mathfrak{C}^{\infty}(\mathbb{R},\mathbb{R}^w)$ . Given two behaviors  $\mathfrak{B}_1$  and  $\mathfrak{B}_2$ , their sum (as subspaces) is denoted by  $\mathfrak{B}_1 + \mathfrak{B}_2$ . If their intersection  $\mathfrak{B}_1 \cap \mathfrak{B}_2 = \{0\}$ , then their sum is called a *direct sum*, and we denote it by  $\mathfrak{B}_1 \oplus \mathfrak{B}_2$ .

We have defined a linear time-invariant differential system as the subspace consisting of all solutions of a set of linear differential equations. In general there are many sets of differential equations leading to one and the same behavior. Any such set of equations is called a *representation* of the behavior. Let  $R \in \mathbb{R}^{g \times w}[\xi]$  be a polynomial matrix. If the behavior  $\mathfrak{B}$  is represented by  $R(\frac{d}{dt})w = 0$  then we call this a *kernel representation* of  $\mathfrak{B}$  and we write  $\mathfrak{B} = \ker(R(\frac{d}{dt}))$ . A kernel representation of  $\mathfrak{B}$  has at least g rows. A given kernel representation,  $R(\frac{d}{dt})w = 0$ , is minimal if and only if the polynomial matrix R has full row rank.

Let  $\mathfrak{B} \in \mathfrak{L}^{\mathsf{w}}$  and let  $R(\frac{\mathrm{d}}{\mathrm{d}t})w = 0$  be a kernel representation of  $\mathfrak{B}$ . Assume rank $(R) < \mathsf{w}$  (which also means that the system is *under-determined*: the number of variables is strictly larger than the number of equations). Then, obviously, some components of  $w = \operatorname{col}(w_1, w_2, \ldots, w_{\mathsf{w}})$  are unconstrained by the requirement  $w \in \mathfrak{B}$ . These components are said to be *free* in  $\mathfrak{B}$ . The maximum number of such components is called the *input cardinality* of  $\mathfrak{B}$  (denoted as  $\mathfrak{m}(\mathfrak{B})$ ). Once  $\mathfrak{m}(\mathfrak{B})$  free components are chosen, the remaining  $\mathsf{w} - \mathfrak{m}(\mathfrak{B})$  components are determined up to a finite-dimensional affine subspace of  $\mathfrak{C}^{\infty}(\mathbb{R}, \mathbb{R}^{\mathsf{w}-\mathfrak{m}(\mathfrak{B})})$ . These are called *outputs*, and the number of outputs is denoted by  $p(\mathfrak{B})$ , called the *output cardinality* of  $\mathfrak{B}$ . Thus, possibly after a permutation of components,  $w \in \mathfrak{B}$  can be partitioned as w = (u, y), with the  $\mathfrak{m}(\mathfrak{B})$  components of uas inputs, and the  $\mathfrak{p}(\mathfrak{B})$  components of y as outputs. We say that (u, y) is an *input/output partition*, in short *i/o partition*, of  $w \in \mathfrak{B}$ , with input u and output y.

The input/output structure of  $\mathfrak{B} \in \mathfrak{L}^{\mathsf{w}}$  is reflected in its kernel representations as follows. Suppose  $R(\frac{\mathrm{d}}{\mathrm{d}t})w = 0$  is a minimal kernel representation of  $\mathfrak{B}$ . Partition  $R = (Q \ P)$ , and accordingly  $w = (w_1, w_2)$ . Then  $w = (w_1, w_2)$  is an i/o partition (with input  $w_1$  and output  $w_2$ ) if and only if P is square and nonsingular.

We now review a number of important properties of behaviors.

- Definition 2.1: 1) A system  $\mathfrak{B} \in \mathfrak{L}^{w}$  is controllable if for all  $w \in \mathfrak{B}$ , there exists a  $T \ge 0$  and a  $w' \in \mathfrak{B}$  such that w(t) = w'(t) for t < 0 and w'(t + T) = 0 for  $t \ge 0$ .
- A system 𝔅 ∈ 𝔅<sup>w</sup> is stabilizable if for all w ∈ 𝔅, there exists a w' ∈ 𝔅 such that w(t) = w'(t) for t < 0 and lim<sub>t→∞</sub> w'(t) = 0.
- A system 𝔅 ∈ 𝔅<sup>w</sup> is anti-stabilizable if for all nonzero w ∈ 𝔅, there exists a w' ∈ 𝔅 such that w(t) = w'(t) for t < 0 and lim<sub>t→∞</sub> w'(t) does not exist.
- 4) A system  $\mathfrak{B} \in \mathfrak{L}^{\mathsf{w}}$  is called *autonomous* if for every  $w \in \mathfrak{B}$  we have that w(t) = 0 for all  $t \leq 0$  implies w = 0.
- 5) A system  $\mathfrak{B} \in \mathfrak{L}^{\mathbb{W}}$  is called *stable* if for every  $w \in \mathfrak{B}$  we have  $\lim_{t\to\infty} w(t) = 0$ , i.e. if all trajectories in the behavior tend to zero as time tends to infinity.
- A system 𝔅 ∈ 𝔅<sup>w</sup> is called *anti-stable* if for all nonzero w ∈ 𝔅 we have lim<sub>t→∞</sub> w(t) does not exist.

It was shown in [9] that if  $\mathfrak{B} = \ker(R(\frac{d}{dt}))$ , then  $\mathfrak{B}$  is autonomous if and only if R has full column rank and is stable if and only if  $R(\lambda)$  has full column rank for all  $\lambda \in \mathbb{C}^+$ , where  $\mathbb{C}^+ = \{\lambda | \operatorname{Re}(\lambda) \ge 0\}$ . Note that a stable behavior is necessarily autonomous, and the same holds for anti-stable behaviors. It can be shown that if  $\mathfrak{B}$  is autonomous and  $\mathfrak{B} = \ker(R(\frac{d}{dt}))$ , then  $\mathfrak{B}$  is anti-stable if and only if  $R(\lambda)$  has full column rank for all  $\lambda \in \mathbb{C}^- = \{\lambda | \operatorname{Re}(\lambda) < 0\}$ . A behavior  $\mathfrak{B} \in \mathfrak{L}^{\mathsf{w}}$  is called *trivial* if  $\mathfrak{B} = \{0\}$ . The trivial behavior is the only behavior that is both controllable and autonomous, as well as both stable and anti-stable. The next proposition states that every autonomous behavior can be written in a unique way as a direct sum of a stable and an anti-stable behavior:

Proposition 2.2: Let  $\mathfrak{B} \in \mathfrak{L}^{w}$  be autonomous. Then there exists a unique stable  $\mathfrak{B}_{-} \in \mathfrak{L}^{w}$ , and a unique anti-stable  $\mathfrak{B}_{+} \in \mathfrak{L}^{w}$  such that  $\mathfrak{B} = \mathfrak{B}_{-} \oplus \mathfrak{B}_{+}$ .

*Proof:* Let R be a full column rank polynomial matrix such that  $\mathfrak{B} = \ker(R(\frac{\mathrm{d}}{\mathrm{d}t}))$ . Let U and V be unimodular matrices such that

$$URV = \begin{pmatrix} D\\ 0 \end{pmatrix}$$

with  $D = \text{diag}(d_1, d_2, \ldots, d_u)$  the Smith form of R. Clearly then  $\mathfrak{B} = V(\frac{d}{dt}) \text{ker}(D(\frac{d}{dt}))$ . Factor each polynomial  $d_i$  as  $d_i^+ d_i^-$  into the product of its anti-stable and stable parts. Let  $D_+$  resp.  $D_-$  be the diagonal matrices with the polynomials  $d_i^+$  resp.  $d_i^-$  on the diagonal. Then  $\text{ker}(D(\frac{d}{dt})) =$  3

 $\ker(D_+(\frac{d}{dt})) \oplus \ker(D_-(\frac{d}{dt}))$  is a unique decomposition into an anti-stable and a stable part (this follows immediately from the corresponding fact for scalar differential equations). Now define  $\mathfrak{B}_+ := V(\frac{d}{dt})\ker(D_+(\frac{d}{dt}))$  and  $\mathfrak{B}_- :=$  $V(\frac{d}{dt})\ker(D_-(\frac{d}{dt}))$ . Then clearly  $\mathfrak{B}_+$  is anti-stable and  $\mathfrak{B}_$ is stable and  $\mathfrak{B} = \mathfrak{B}_+ \oplus \mathfrak{B}_-$ . Uniqueness follows simply from the uniqueness of the decomposition  $\ker(D(\frac{d}{dt})) =$  $\ker(D_+(\frac{d}{dt})) \oplus \ker(D_-(\frac{d}{dt}))$  into anti-stable and stable parts.

The behaviors  $\mathfrak{B}_{-}$  and  $\mathfrak{B}_{+}$  are called the *stable part* and the *anti-stable part* of  $\mathfrak{B}$ , respectively. It follows from Theorem 3.2.16 in [9] that all trajectories in  $\mathfrak{B}_{-}$  are stable Bohl functions and all nonzero trajectories in  $\mathfrak{B}_{+}$  are anti-stable Bohl functions.

It was shown in [9] that any  $\mathfrak{B} \in \mathfrak{L}^{w}$  contains a largest controllable subbehavior. This behavior is called the *controllable part* of the behavior  $\mathfrak{B}$  and is denoted by  $\mathfrak{B}_{cont}$ . In a similar way it can be shown that any  $\mathfrak{B}$  contains a largest stabilizable subbehavior and a largest anti-stabilizable sub behavior. We substantiate this in the following theorem:

Theorem 2.3: Let  $\mathfrak{B} \in \mathfrak{L}^{w}$ . There exists a largest stabilizable subbehavior contained in  $\mathfrak{B}$ , denoted by  $\mathfrak{B}_{stab}$  and called the *stabilizable part* of  $\mathfrak{B}$ . Likewise there exists a largest anti-stabilizable subbehavior of  $\mathfrak{B}$ , denoted by  $\mathfrak{B}_{antistab}$  and called the *anti-stabilizable part* of  $\mathfrak{B}$ . Their intersection is equal to the controllable part:  $\mathfrak{B}_{cont} = \mathfrak{B}_{stab} \cap \mathfrak{B}_{antistab}$ .

**Proof:** Let  $R(\frac{d}{dt})w = 0$  be a minimal kernel representation of  $\mathfrak{B}$ . Let U and V be unimodular matrices such that  $URV = (D \ 0)$  where  $(D \ 0)$  is the Smith form of R,  $D = \text{diag}(d_1, d_2, \ldots, d_p)$ . It was shown in [9] that  $\mathfrak{B}_{\text{cont}} = V(\frac{d}{dt}) \text{ker}((I_p \ 0))$ . Factor each polynomial  $d_i$  as  $d_i^+ d_i^-$  into the product of its anti-stable and stable parts. Let  $D_+$  resp.  $D_-$  be the diagonal matrices with the  $d_i^+$  resp.  $d_i^-$  on the diagonal. Define the following subbehaviors of  $\mathfrak{B}$ :

$$\mathfrak{B}_{\text{stab}} := V(\frac{\mathrm{d}}{\mathrm{d}t}) \operatorname{ker}(\left(D_{-}(\frac{\mathrm{d}}{\mathrm{d}t}) \quad 0\right)),$$
  
$$\mathfrak{B}_{\text{antistab}} := V(\frac{\mathrm{d}}{\mathrm{d}t}) \operatorname{ker}(\left(D_{+}(\frac{\mathrm{d}}{\mathrm{d}t}) \quad 0\right)).$$
 (1)

It follows immediately from Definition 2.1 that  $\mathfrak{B}_{stab}$  is stabilizable and  $\mathfrak{B}_{\mathrm{antistab}}$  is anti-stabilizable. We will now show that they are the largest stabilizable and anti-stabilizable subbehaviors of  $\mathfrak{B}$ , respectively. Let  $\mathfrak{B}_1$  be a stabilizable subbehavior of  $\mathfrak{B}$ , and let  $R_1(\frac{\mathrm{d}}{\mathrm{d}t})w = 0$  be a minimal kernel representation. Let  $U_1$  and  $V_1$  be unimodular matrices such that  $U_1R_1V_1 = (D_1 \quad 0)$  with  $D_1$  the Smith form of  $R_1$ . Let  $\overline{R}_1 := (D_1 \ 0)V_1^{-1}$  and  $\overline{R} := (D_- \ 0)V^{-1}$ . Then  $\mathfrak{B}_1 = \ker(\overline{R}_1(\frac{\mathrm{d}}{\mathrm{d}t})), \mathfrak{B}_{\mathrm{stab}} = \ker(\overline{R}(\frac{\mathrm{d}}{\mathrm{d}t}))$  and  $\mathfrak{B} = \ker(D_+(\frac{\mathrm{d}}{\mathrm{d}t})\overline{R}(\frac{\mathrm{d}}{\mathrm{d}t}))$ . Since  $\mathfrak{B}_1 \subseteq \mathfrak{B}$ , there exists a polynomial matrix F such that  $D_+\overline{R} = F\overline{R}_1$ . By inspection, there exists a stable rational matrix T such that  $\overline{R}_1T = I$ . Thus we obtain  $\overline{R}T = D_{+}^{-1}F$ . The left hand side is stable, while the right hand side is anti-stable. Thus  $D_{+}^{-1}F$  must be a polynomial matrix. It then follows from  $\overline{R} = D_{+}^{-1} F \overline{R}_1$  that  $\mathfrak{B}_1 \subset \mathfrak{B}_{stab}$ . A similar proof applies to  $\mathfrak{B}_{antistab}$ . Finally, it can be verified immediately that  $\mathfrak{B}_{stab} \cap \mathfrak{B}_{antistab} = \mathfrak{B}_{cont}$ .

It was also shown in [9] that a given  $\mathfrak{B} \in \mathfrak{L}^{w}$  can always be decomposed as  $\mathfrak{B} = \mathfrak{B}_{cont} \oplus \mathfrak{B}_{aut}$ , where  $\mathfrak{B}_{cont}$  is the (unique) controllable part of  $\mathfrak{B}$ , and  $\mathfrak{B}_{aut}$  is a (nonunique) autonomous subbehavior of  $\mathfrak{B}$ . In fact, using the notation in the proof of Theorem 2.3, it was shown in [9] that the autonomous subbehavior

$$\overline{\mathfrak{B}}_{\mathrm{aut}} := V(\frac{\mathrm{d}}{\mathrm{d}t}) \mathrm{ker}(\begin{pmatrix} D(\frac{\mathrm{d}}{\mathrm{d}t}) & 0\\ 0 & I \end{pmatrix})$$

satisfies  $\mathfrak{B} = \mathfrak{B}_{cont} \oplus \overline{\mathfrak{B}}_{aut}$ . Clearly, its stable part is equal to

$$(\overline{\mathfrak{B}}_{\mathrm{aut}})_{-} = V(\frac{\mathrm{d}}{\mathrm{d}t})\ker(\begin{pmatrix} D_{-}(\frac{\mathrm{d}}{\mathrm{d}t}) & 0\\ 0 & I \end{pmatrix}),$$

and similarly its anti-stable part  $(\overline{\mathfrak{B}}_{aut})_+$  has this representation with  $D_-$  replaced by  $D_+$ . It then immediately follows from (1) that  $\mathfrak{B}_{stab} = \mathfrak{B}_{cont} \oplus (\overline{\mathfrak{B}}_{aut})_-$  and  $\mathfrak{B}_{antistab} = \mathfrak{B}_{cont} \oplus (\overline{\mathfrak{B}}_{aut})_+$ . It is a remarkable fact that, a fortiori, for any choice of autonomous complement  $\mathfrak{B}_{aut}$  the stabilizable part of  $\mathfrak{B}$  is equal to the direct sum of the controllable part of  $\mathfrak{B}$  and the stable part of  $\mathfrak{B}_{aut}$ , and the anti-stabilizable part of  $\mathfrak{B}$  is equal to the direct sum of the controllable part of  $\mathfrak{B}$  and the anti-stable part of  $\mathfrak{B}_{aut}$ :

 $\begin{array}{l} \textit{Proposition 2.4: Let } \mathfrak{B} \in \mathfrak{L}^{w}. \ \textit{Let } \mathfrak{B}_{aut} \ \textit{be any autonomous} \\ \textit{behavior such that } \mathfrak{B} = \mathfrak{B}_{cont} \oplus \mathfrak{B}_{aut}. \ \textit{Then we have } \mathfrak{B}_{stab} = \\ \mathfrak{B}_{cont} \oplus (\mathfrak{B}_{aut})_{-} \ \textit{and } \mathfrak{B}_{antistab} = \mathfrak{B}_{cont} \oplus (\mathfrak{B}_{aut})_{+}. \end{array}$ 

**Proof:** Let  $\mathfrak{B}_{aut}$  be any autonomous direct summand of  $\mathfrak{B}_{cont}$  as in the statement of the proposition. Using the notation in the proof of Theorem 2.3 and the result of Exercise 5.6, page 190 in [9] (see also [10], Lemma 4.2) it can be shown that there exists a polynomial matrix F and a unimodular matrix S such that  $\mathfrak{B}_{aut}$  is represented as

$$\mathfrak{B}_{\rm aut} = V(\frac{\rm d}{{\rm d}t}) {\rm ker}(\begin{pmatrix} D(\frac{\rm d}{{\rm d}t}) & 0\\ F(\frac{\rm d}{{\rm d}t}) & S(\frac{\rm d}{{\rm d}t}) \end{pmatrix})$$

Using uniqueness in Theorem 2.2 it is then easily verified that

$$(\mathfrak{B}_{\text{aut}})_{-} = V(\frac{\mathrm{d}}{\mathrm{d}t}) \operatorname{ker} \begin{pmatrix} D_{-}(\frac{\mathrm{d}}{\mathrm{d}t}) & 0\\ F(\frac{\mathrm{d}}{\mathrm{d}t}) & S(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix}$$

A similar representation holds for  $(\mathfrak{B}_{aut})_+$  with  $D_-$  replaced by  $D_+$ . It is then a matter of straightforward verification to check that  $\mathfrak{B}_{stab} = \mathfrak{B}_{cont} \oplus (\mathfrak{B}_{aut})_-$  and  $\mathfrak{B}_{antistab} = \mathfrak{B}_{cont} \oplus (\mathfrak{B}_{aut})_+$ .

It follows immediately from this that, for a given autonomous  $\mathfrak{B} \in \mathfrak{L}^{w}$ , its stable part  $\mathfrak{B}_{-}$  is in fact equal to the stabilizable part  $\mathfrak{B}_{stab}$  of  $\mathfrak{B}$ , and the anti-stable part  $\mathfrak{B}_{+}$  is equal to the anti-stabilizable part  $\mathfrak{B}_{antistab}$  of  $\mathfrak{B}$ . For this reason, in the sequel we will allow ourselves to denote  $\mathfrak{B}_{-}$  by  $\mathfrak{B}_{stab}$  and  $\mathfrak{B}_{+}$  by  $\mathfrak{B}_{antistab}$  for given *autonomous*  $\mathfrak{B}$ . Also note that  $\mathfrak{B}$  is autonomous if and only if  $\mathfrak{B}_{cont} = \{0\}$ .

It was shown in [9] that controllable behaviors are exactly those that admit an image representation. To be precise,  $\mathfrak{B}$ is controllable if and only if there exists a  $w \times l$  polynomial matrix M such that

$$\mathfrak{B} = \{ M(\frac{\mathrm{d}}{\mathrm{d}t})\ell \mid \ell \in \mathfrak{C}^{\infty}(\mathbb{R}, \mathbb{R}^l) \}.$$

This representation of  $\mathfrak{B}$  is called an *image representation*, and we write  $\mathfrak{B} = \operatorname{im}(M(\frac{\mathrm{d}}{\mathrm{d}t}))$ .

To conclude this section, we review some facts on elimination of variables. Let  $\mathfrak{B} \in \mathcal{L}^{w_1+w_2}$  with system variable  $w = (w_1, w_2)$ . Let  $P_{w_1}$  denote the projection onto the  $w_1$ component. Then the set  $P_{w_1}\mathfrak{B}$ , consisting of all  $w_1$  for which there exists  $w_2$  such that  $(w_1, w_2) \in \mathfrak{B}$ , is again a linear timeinvariant differential system. We denote  $P_{w_1}\mathfrak{B}$  by  $\mathfrak{B}_{w_1}$ , and call it the behavior obtained by eliminating  $w_2$  from  $\mathfrak{B}$ , or the projection of  $\mathfrak{B}$  onto  $w_1$ .

If  $\mathfrak{B} = \ker \left( R_1(\frac{d}{dt}) \quad R_2(\frac{d}{dt}) \right)$  then a representation for  $\mathfrak{B}_{w_1}$  is obtained as follows: choose a unimodular matrix U such that

$$UR_2 = \begin{pmatrix} R_{12} \\ 0 \end{pmatrix},$$

with  $R_{12}$  full row rank, and conformably partition

$$UR_1 = \begin{pmatrix} R_{11} \\ R_{21} \end{pmatrix}.$$

Then  $\mathfrak{B}_{w_1} = \ker(R_{21}(\frac{\mathrm{d}}{\mathrm{d}t}))$  (see [9], Section 6.2.2).

For linear time-invariant differential systems it can be shown that the two operations of taking the controllable part and projecting onto a variable commute, i.e.  $(\mathfrak{B}_{\text{cont}})_{w_1} = (\mathfrak{B}_{w_1})_{\text{cont}}$ for all  $\mathfrak{B} \in \mathfrak{L}^{\mathfrak{w}_1 + \mathfrak{w}_2}$  (see e.g. Lemma 2.10.4 in [11]).

An important role is also played by the behavior obtained from  $\mathfrak{B} \in \mathfrak{L}^{\mathfrak{w}_1+\mathfrak{w}_2}$  by requiring  $w_1 = 0$ . This behavior is denoted by  $\mathcal{N}_{w_2}(\mathfrak{B})$ , and is defined as

$$\mathcal{N}_{w_2}(\mathfrak{B}) = \{ w_2 \mid (0, w_2) \in \mathfrak{B} \},\$$

called the *hidden behavior* of  $w_2$  in  $\mathfrak{B}$ . If  $\mathfrak{B} = \ker \left(R_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \quad R_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\right)$  then  $\mathcal{N}_{w_2}(\mathfrak{B}) = \ker R_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)$ .

## III. ACHIEVABILITY

Definition 3.1: Given a linear time-invariant differential system  $(\mathbb{R}, \mathbb{R}^{w_1+w_2}, \mathcal{P})$ , the *plant*, and another linear time-invariant differential system  $(\mathbb{R}, \mathbb{R}^{w_1+w_2}, \mathcal{O})$ , we call the partial interconnection  $\mathcal{P} \wedge_{w_1} \mathcal{O}$  of  $\mathcal{P}$  and  $\mathcal{O}$  through  $w_1$  an observer interconnection, and  $\mathcal{O}$  an observer for  $w_2$  from  $w_1$  (in  $\mathcal{P}$ ).

To avoid confusion we usually label the second set of variables in  $\mathcal{P}$  by  $w_2$  while we label the second set of variables in  $\mathcal{O}$  by  $\hat{w}_2$ . The first set of variables in both  $\mathcal{P}$  and  $\mathcal{O}$  is labelled by  $w_1$  since it is shared in the observer interconnection. Given an observer interconnection,  $\hat{w}_2$  is interpreted as an estimate for  $w_2$ . This makes sense since they are both of dimension  $w_2$ . Note, though, that the arrangement in an observer interconnection is completely symmetric, and hence we could equally well think of  $\mathcal{P}$  as an "observer" for  $\hat{w}_2$  from  $w_1$  (in  $\mathcal{O}$ ). Choosing to call  $\mathcal{P}$  the plant and  $\mathcal{O}$  the observer merely indicates our preferred interpretation.



Fig. 1. An observer interconnection gives rise to an error behavior through interconnection with the "differencing system"  $\mathcal{D}$ .

Definition 3.2: [2] Given an observer interconnection, the associated error behavior  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  is defined as

$$\mathcal{E}(\mathcal{P}, \mathcal{O}) = \left( \left( \mathcal{P} \wedge_{w_1} \mathcal{O} \right) \wedge_{(w_2, \hat{w}_2)} \mathcal{D} \right)_e \\ = \left\{ e = \hat{w}_2 - w_2 \, | \, \exists_{w_1}(w_1, w_2) \in \mathcal{P}, (w_1, \hat{w}_2) \in \mathcal{O} \right\}$$

where  $\mathcal{D} = \{(w_2, \hat{w}_2, e) | e = \hat{w}_2 - w_2\}$  defines the *error* variable *e*. The dynamical system  $(\mathbb{R}, \mathbb{R}^{w_2}, \mathcal{E}(\mathcal{P}, \mathcal{O}))$  is then also called the associated *error system*.

The total interconnection giving rise to the error behavior is depicted in Figure 1. The error behavior is the projection of this total interconnection onto the variable e. Note that this notion is still perfectly symmetric with respect to interchanging the roles of the plant  $\mathcal{P}$  and the observer  $\mathcal{O}$ , except for the sign of the variable e.

Definition 3.3: Given a plant  $\mathcal{P} \in \mathfrak{L}^{w_1+w_2}$ , a behavior  $\mathcal{E} \in \mathfrak{L}^{w_2}$  is an *achievable error behavior* (for  $\mathcal{P}$ ) if there exists an observer  $\mathcal{O}$  for  $w_2$  from  $w_1$  (in  $\mathcal{P}$ ) such that  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \mathcal{E}$ .

We can characterize all achievable error behaviors for a given plant  $\mathcal{P}$  in terms of the hidden behavior  $\mathcal{N}_{w_2}(\mathcal{P})$  of  $w_2$  in  $\mathcal{P}$ . This is the content of Proposition 3.5 below. Its proof uses the following lemma.

*Lemma 3.4:* Given an observer interconnection, let  $\mathcal{P} = \ker \left(R_1(\frac{\mathrm{d}}{\mathrm{d}t}) \quad R_2(\frac{\mathrm{d}}{\mathrm{d}t})\right)$  be a minimal kernel representation and let  $\mathcal{O} = \ker \left(S(\frac{\mathrm{d}}{\mathrm{d}t})R_1(\frac{\mathrm{d}}{\mathrm{d}t}) \quad S(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t})\right)$ , where S is a polynomial matrix. Then  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \ker \left(S(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t})\right)$ .

*Proof:*  $(\mathcal{P} \wedge_{w_1} \mathcal{O}) \wedge_{(w_2, \hat{w}_2)} \mathcal{D}$  is given by the equation

$$\begin{pmatrix} R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & R_2(\frac{\mathrm{d}}{\mathrm{d}t}) & 0\\ 0 & I & -I\\ S(\frac{\mathrm{d}}{\mathrm{d}t})R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & 0 & S(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix} \begin{pmatrix} w_1\\ w_2\\ \hat{w}_2 \end{pmatrix} = \begin{pmatrix} 0\\ I\\ 0 \end{pmatrix} e_1$$

Using unimodular row transformations this can be equivalently expressed as

$$\begin{pmatrix} R_1(\frac{d}{dt}) & R_2(\frac{d}{dt}) & 0\\ 0 & I & -I\\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} w_1\\ w_2\\ \hat{w}_2 \end{pmatrix} = \begin{pmatrix} 0\\ I\\ S(\frac{d}{dt})R_2(\frac{d}{dt}) \end{pmatrix} e,$$

where in the matrix on the left the submatrix consisting of the first two block rows has full row rank. But then a kernel representation for the projected behavior  $((\mathcal{P} \wedge_{w_1} \mathcal{O}) \wedge_{(w_2, \hat{w}_2)} \mathcal{D})_e$  is given by the third block row on the right, i.e.  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \ker(S(\frac{d}{dt})R_2(\frac{d}{dt})).$ 

Proposition 3.5: Given an observer interconnection then  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \mathcal{N}_{w_2}(\mathcal{P} + \mathcal{O})$ , where '+' denotes the sum of linear subspaces of  $\mathfrak{C}^{\infty}(\mathbb{R}, \mathbb{R}^{w_1+w_2})$ . Hence,  $\mathcal{E}$  is an achievable error behavior (for  $\mathcal{P}$ ) if and only if  $\mathcal{N}_{w_2}(\mathcal{P}) \subset \mathcal{E}$ , i.e. if and only if it contains the hidden behavior of  $w_2$  in  $\mathcal{P}$ .

*Proof:* Consider an observer interconnection and  $e \in \mathcal{E}(\mathcal{P}, \mathcal{O})$ . Then there exist  $(w_1, w_2) \in \mathcal{P}$  and  $(w_1, \hat{w}_2) \in \mathcal{O}$  such that  $e = \hat{w}_2 - w_2$ , i.e.  $(0, e) = (-w_1 + w_1, -w_2 + \hat{w}_2) \in \mathcal{P} + \mathcal{O}$ . Hence  $e \in \mathcal{N}_{w_2}(\mathcal{P} + \mathcal{O})$ . Conversely, let  $e \in \mathcal{N}_{w_2}(\mathcal{P} + \mathcal{O})$  then  $(0, e) \in \mathcal{P} + \mathcal{O}$ . But then there exist  $(\tilde{w}_1, \tilde{w}_2) \in \mathcal{P}$  and  $(w_1, \hat{w}_2) \in \mathcal{O}$  with  $(0, e) = (\tilde{w}_1 + w_1, \tilde{w}_2 + \hat{w}_2)$ . Hence  $w_1 = -\tilde{w}_1$  and by linearity  $(w_1, w_2) \in \mathcal{P}$  where  $w_2 := -\tilde{w}_2$ . We conclude  $e = \hat{w}_2 - w_2 \in \mathcal{E}(\mathcal{P}, \mathcal{O})$ . This proves the formula for  $\mathcal{E}(\mathcal{P}, \mathcal{O})$ .

Assume now that  $\mathcal{E}$  is achieved by  $\mathcal{O}$  then  $\mathcal{E} = \mathcal{E}(\mathcal{P}, \mathcal{O}) = \mathcal{N}_{w_2}(\mathcal{P} + \mathcal{O}) \supset \mathcal{N}_{w_2}(\mathcal{P}).$ 

Conversely, assume that  $\mathcal{N}_{w_2}(\mathcal{P}) \subset \mathcal{E}$ . Let  $\mathcal{P} = \ker\left(R_1(\frac{\mathrm{d}}{\mathrm{d}t}) \ R_2(\frac{\mathrm{d}}{\mathrm{d}t})\right)$  be a minimal kernel representation then  $\mathcal{N}_{w_2}(\mathcal{P}) = \ker(R_2(\frac{\mathrm{d}}{\mathrm{d}t}))$ . Let  $\mathcal{E} = \ker(E(\frac{\mathrm{d}}{\mathrm{d}t}))$  then there exists a polynomial matrix S such that  $E = SR_2$ . Define  $\mathcal{O} = \ker\left(S(\frac{\mathrm{d}}{\mathrm{d}t})R_1(\frac{\mathrm{d}}{\mathrm{d}t}) \ S(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t})\right)$ . By Lemma 3.4  $\mathcal{E}(\mathcal{P},\mathcal{O}) = \ker(S(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t})) = \ker(E(\frac{\mathrm{d}}{\mathrm{d}t})) = \mathcal{E}$ , and hence  $\mathcal{O}$  achieves  $\mathcal{E}$ .

Note that the second part of the above proof is constructive. Given any achievable error behavior  $\mathcal{E}$ , it uses kernel representations to explicitly construct an observer that achieves  $\mathcal{E}$ . By construction, this observer contains the plant behavior,  $\mathcal{P} \subset \mathcal{O}$ . In previous work [3], such observers have been called *consistent*.

*Remark 3.6:* Owing to the symmetry between the plant and the observer in an observer interconnection, the associated error behavior will always contain the hidden behavior  $\mathcal{N}_{\hat{w}_2}(\mathcal{O})$  of  $\hat{w}_2$  in  $\mathcal{O}$ . This is also apparent from the formula for  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  in Proposition 3.5.

*Remark 3.7:* The above set-up can be generalized so that the plant  $\mathcal{P}$ , in addition to  $w_1$  and  $w_2$ , has a third variable, say  $w_3$ , representing, for example, an unknown disturbance. In



Fig. 2. An observer interconnection with plant disturbance  $w_3$ .

that case, the plant is a system  $(\mathbb{R}, \mathbb{R}^{w_1+w_2+w_3}, \mathcal{P})$  with variable  $(w_1, w_2, w_3)$ , and the error behavior  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  resulting from the observer interconnection of the plant and the observer  $(\mathbb{R}, \mathbb{R}^{w_1+w_2}, \mathcal{O})$  is given by

$$\mathcal{E}(\mathcal{P}, \mathcal{O}) = \left( \left( \mathcal{P} \wedge_{w_1} \mathcal{O} \right) \wedge_{(w_2, \hat{w}_2)} \mathcal{D} \right)_e \\ = \left\{ e = \hat{w}_2 - w_2 \, | \, \exists_{w_1, w_3}(w_1, w_2, w_3) \in \mathcal{P}, \\ (w_1, \hat{w}_2) \in \mathcal{O} \right\}$$

It is immediately clear that, in fact,  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \mathcal{E}(\mathcal{P}_{(w_1,w_2)}, \mathcal{O})$ , in other words the error behavior is equal to the error behavior associated with the projected plant behavior  $\mathcal{P}_{(w_1,w_2)}$  and the observer  $\mathcal{O}$ . Using this observation, all results of this paper can immediately be generalized by applying them to the projected plant behavior after elimination of the additional variable  $w_3$ . Therefore, in the sequel, the main text will deal with the set-up where the plant has variable  $(w_1, w_2)$ . Where appropriate, we will comment on the extension to additional variables. We also refer to Section VII for an application of this technique of elimination.

#### **IV. EXISTENCE**

Given a plant, existence results for observers are typically associated with particular, desirable properties of the observer and/or the resulting error system. For example, one could ask whether there exists an observer with a stable associated error behavior.

It is clear that Proposition 3.5 immediately translates into general existence results regarding properties of the error behavior that are hereditary with repect to behavior inclusion. For example, any subbehavior of an autonomous (stable, trivial) behavior is also autonomous (stable, trivial), and hence the existence of observers with an autonomous (stable, trivial) associated error behavior depends solely on the respective properties of the hidden behavior  $\mathcal{N}_{w_2}(\mathcal{P})$ , i.e. on an associated property of the observed plant. We first recall the definitions of these plant properties, cf. [2], [12].

Definition 4.1: Given a linear time-invariant differential system  $(\mathbb{R}, \mathbb{R}^{w_1+w_2}, \mathcal{P})$ , the variable  $w_2$  is

- (1) observable from w<sub>1</sub> (in P) if for all (w<sub>1</sub>, w<sub>2</sub>) ∈ P, w<sub>1</sub> = 0 implies w<sub>2</sub> = 0, i.e. if N<sub>w<sub>2</sub></sub>(P) = {0}.
- (2) detectable from w₁ (in P) if for all (w₁, w₂) ∈ P, w₁ = 0 implies lim<sub>t→∞</sub> w₂(t) = 0, i.e. if N<sub>w₂</sub>(P) is stable.
- (3) trackable from  $w_1$  (in  $\mathcal{P}$ ) if for all  $(w_1, w_2) \in \mathcal{P}$ ,  $w_1 = 0$ and  $w_2(t) = 0$  for all  $t \leq 0$  implies  $w_2 = 0$ , i.e. if  $\mathcal{N}_{w_2}(\mathcal{P})$  is autonomous.

Clearly, observable implies detectable which in turn implies trackable.

*Remark 4.2:* For the special case where  $w_2$  is an output and  $w_1$  is the corresponding input,  $\mathcal{N}_{w_2}(\mathcal{P})$  is autonomous so  $w_2$  is automatically trackable from  $w_1$  (in  $\mathcal{P}$ ). This includes the case where  $w_2$  contains state variables.

Usually, the dynamic properties of an error behavior associated with a plant and an observer are attributed to the observer since the plant is thought of as given. We recall some of these properties, cf. [2], [12].

*Definition 4.3:* Given an observer interconnection, the observer is

(1) *exact* if  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \{0\}$ , i.e. if e = 0.

- (2) asymptotic if  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  is stable, i.e. if  $\lim_{t\to\infty} e(t) = 0$ .
- (3) tracking if  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  is autonomous, i.e. if e(t) = 0 for all  $t \leq 0$  implies e = 0.

The following existence results (cf. [2], [12]) are now immediate consequences of these definitions and Proposition 3.5. *Proposition 4.4:* Let  $\mathcal{P} \in \mathfrak{L}^{w_1+w_2}$  be given.

- (1) There exists an exact observer for  $w_2$  from  $w_1$  (in  $\mathcal{P}$ ) if and only if  $w_2$  is observable from  $w_1$  (in  $\mathcal{P}$ ).
- (2) There exists an asymptotic observer for  $w_2$  from  $w_1$  (in  $\mathcal{P}$ ) if and only if  $w_2$  is detectable from  $w_1$  (in  $\mathcal{P}$ ).
- (3) There exists a tracking observer for  $w_2$  from  $w_1$  (in  $\mathcal{P}$ ) if and only if  $w_2$  is trackable from  $w_1$  (in  $\mathcal{P}$ ).

*Remark 4.5:* In line with Remark 3.7 we note that the statements of Prop. 4.4 remain valid if the plant  $\mathcal{P}$  has an additional variable  $w_3$ . This does however require a modification of the definitions of observability, detectability and trackability to account for the presence of  $w_3$ , cf. [2]. Given a linear time-invariant differential system ( $\mathbb{R}, \mathbb{R}^{w_1+w_2+w_3}, \mathcal{P}$ ), the variable  $w_2$  is called *observable from*  $w_1$  (*in*  $\mathcal{P}$ ) if for

all  $(w_1, w_2, w_3) \in \mathcal{P}$ ,  $w_1 = 0$  implies  $w_2 = 0$ . It is easily verified that this condition holds if and only if in the projected behavior  $\mathcal{P}_{(w_1, w_2)}$  the variable  $w_2$  is observable from  $w_1$ . The fact that Prop. 4.4 remains valid is a simple consequence of this. Likewise we can modify the definitions of detectability and trackability. The details are left to the reader. We note that in the state space context, where  $w_3$  is usually interpreted as an unknown input,  $w_1$  as the observed output, and  $w_2$  as the to be estimated variable, the above properties are often called *strong* observability and *strong* detectability of  $w_2$  from  $w_1$ , respectively [13].

We will dwell a little bit on just how general our notion of an observer is, e.g. compared to the notion of an observer as defined in [2]. In an observer interconnection as defined above, the observer can impose restrictions on the variable  $w_1$  that are not already present in the plant. For example, the observer can impose the equation  $w_1 = 0$ . It is a matter of taste whether one wishes to call such a system an "observer", since it "interferes" with the operation of the plant. We opt to use the term observer in the broad sense and introduce the following observer property to distinguish observers that do not interfere with the operation of the plant in this way.

Definition 4.6: Given an observer interconnection, the observer is called *nonintrusive* if  $(\mathcal{P} \wedge_{w_1} \mathcal{O})_{(w_1,w_2)} = \mathcal{P}$ , i.e. if the plant behavior is not changed by the observer interconnection.

Clearly, an observer interconnection is nonintrusive if and only if for all  $(w_1, w_2) \in \mathcal{P}$  there exists a  $\hat{w}_2$  such that  $(w_1, w_2, \hat{w}_2) \in \mathcal{P} \wedge_{w_1} \mathcal{O}$  or, equivalently, such that  $(w_1, \hat{w}_2) \in \mathcal{O}$ . In previous work, a nonintrusive observer has also been called an "acceptor" [2]. This notion is slightly weaker than the requirement that the variable  $w_1$  be *free in*  $\mathcal{O}$ , or even that  $w_1$  be an *input in*  $\mathcal{O}$  (with  $\hat{w}_2$  the associated output). The latter type of observers are commmonly called *i/o-observers* and have been studied comprehensively in [14], [15]. For *i/o*observers the question of properness of the observer (i.e. of its associated transfer function) arises, and connections to the classical state observer theory can be drawn.

It is a curious consequence of Proposition 3.5 that an achievable autonomous error behavior is always also achievable with an i/o-observer. In this case, the i/o-structure of the observer can be assumed without loss of generality. More precisely, we have the following result.

Proposition 4.7: Given a plant  $\mathcal{P} \in \mathfrak{L}^{w_1+w_2}$  and an achievable autonomous error behavior  $\mathcal{E}$ , then there exists an i/o-observer  $\mathcal{O}$  for  $w_2$  from  $w_1$  (in  $\mathcal{P}$ ) such that  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \mathcal{E}$ .

**Proof:** Let  $\mathcal{P} = \ker \left(R_1(\frac{d}{dt}) \quad R_2(\frac{d}{dt})\right)$  be a minimal kernel representation. Then  $\mathcal{N}_{w_2}(\mathcal{P}) = \ker \left(R_2(\frac{d}{dt})\right)$ . By Proposition 3.5 there exists a polynomial matrix S such that  $\mathcal{E} = \ker \left(S(\frac{d}{dt})R_2(\frac{d}{dt})\right)$ . Since  $\mathcal{E}$  is autonomous,  $SR_2$  has full column rank and hence there exists a unimodular matrix U such that

$$USR_2 = \begin{pmatrix} R\\0 \end{pmatrix},\tag{2}$$

where  $\bar{R}$  is square and nonsingular. It follows that  $\mathcal{E}$  =

 $\ker(\bar{R}(\frac{\mathrm{d}}{\mathrm{d}t}))$ . Define

$$\binom{S_1}{S_2} = US,$$

where the splitting is as in (2), then  $S_1R_2 = \bar{R}$ . Define  $\mathcal{O} = \ker \left(S_1(\frac{d}{dt})R_1(\frac{d}{dt}) \quad S_1(\frac{d}{dt})R_2(\frac{d}{dt})\right)$ . By Lemma 3.4,  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \ker(S_1(\frac{d}{dt})R_2(\frac{d}{dt})) = \ker(\bar{R}(\frac{d}{dt})) = \mathcal{E}$ .

The previous proposition implies that we can augment any of the statements in Proposition 4.4 to require the existence of an i/o-observer. Note that this does not necessarily imply the existence of a *proper* i/o-observer, making the result maybe a little bit less surprising.

*Remark 4.8:* A further consequence of the symmetry between the plant and the observer in an observer interconnection is that the variable  $\hat{w}_2$  in an exact (stable, tracking) observer will necessarily be observable (detectable, trackable) from  $w_1$ .

In the next section we will derive a fundamental structure theorem for nonintrusive observers.

### V. AN INTERNAL MODEL PRINCIPLE

In this section we will show that every nonintrusive observer giving rise to a "reasonable" error behavior must contain a sizeable part of the plant behavior, i.e. an *internal model* of (part of) the plant dynamics. We begin with two technical lemmas.

Lemma 5.1: Let  $\mathcal{B}_1 \in \mathfrak{L}^{w_1}$  be an autonomous behavior and let  $\mathcal{B}_2 \in \mathfrak{L}^{w_1+w_2}$  be such that  $w_2$  is trackable from  $w_1$  (in  $\mathcal{B}_2$ ). Then  $(\mathcal{B}_1 \wedge_{w_1} \mathcal{B}_2)_{w_2}$  is autonomous. *Proof:* Since  $\mathcal{B}_1$  is autonomous, it admits a kernel rep-

*Proof:* Since  $\mathcal{B}_1$  is autonomous, it admits a kernel representation  $\mathcal{B}_1 = \ker(R(\frac{d}{dt}))$  where R has full column rank.  $w_2$  being trackable from  $w_1$  (in  $\mathcal{B}_2$ ) implies that  $\mathcal{N}_{w_2}(\mathcal{B}_2)$  is autonomous and hence that  $\mathcal{B}_2$  admits a kernel representation  $\mathcal{B}_2 = \ker\left(R_1(\frac{d}{dt}) \quad R_2(\frac{d}{dt})\right)$  where  $R_2$  has full column rank. But then

$$\mathcal{B}_1 \wedge_{w_1} \mathcal{B}_2 = \ker \begin{pmatrix} R(\frac{\mathrm{d}}{\mathrm{d}t}) & 0\\ R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & R_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix}$$

and the matrix on the right has full column rank. Hence  $\mathcal{B}_1 \wedge_{w_1} \mathcal{B}_2$  is autonomous and so is its projection.

*Lemma 5.2:* Consider an observer interconnection where the observer O is nonintrusive. Let

$$\mathcal{P}_{\text{cont}} = \operatorname{im} \begin{pmatrix} M_1(\frac{\mathrm{d}}{\mathrm{d}t}) \\ M_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix} \text{ and } \mathcal{O}_{\text{cont}} = \operatorname{im} \begin{pmatrix} L_1(\frac{\mathrm{d}}{\mathrm{d}t}) \\ L_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix}$$

be image representations of the controllable parts of the plant behavior and of the observer behavior, respectively. Then there exists a rational matrix S such that  $M_1 = L_1S$ . Moreover, if  $\hat{w}_2$  is trackable from  $w_1$  (in  $\mathcal{O}$ ) then there exists a rational matrix X such that  $L_2 = XL_1$ .

*Proof:* By definition of nonintrusiveness, for every  $(w_1, w_2) \in \mathcal{P}$  there exists  $\hat{w}_2$  such that  $(w_1, \hat{w}_2) \in \mathcal{O}$ . This is equivalent to the inclusion of projected behaviors  $\mathcal{P}_{w_1} \subset \mathcal{O}_{w_1}$ . But then also  $(\mathcal{P}_{w_1})_{\text{cont}} \subset (\mathcal{O}_{w_1})_{\text{cont}}$  and hence  $\operatorname{im}(M_1(\frac{\mathrm{d}}{\mathrm{d}t})) = (\mathcal{P}_{\text{cont}})_{w_1} = (\mathcal{P}_{w_1})_{\text{cont}} \subset (\mathcal{O}_{w_1})_{\text{cont}} = (\mathcal{O}_{\text{cont}})_{w_1} = \operatorname{im}(L_1(\frac{\mathrm{d}}{\mathrm{d}t}))$ . The first statement now follows as in the proof of Theorem 7.3 in [10].

Now let  $\hat{w}_2$  be trackable from  $w_1$  (in  $\mathcal{O}$ ) then  $\mathcal{N}_{\hat{w}_2}(\mathcal{O})$  is autonomous. Since  $\mathcal{N}_{\hat{w}_2}(\mathcal{O}_{cont}) \subset \mathcal{N}_{\hat{w}_2}(\mathcal{O})$ , this implies that

 $\mathcal{N}_{\hat{w}_2}(\mathcal{O}_{\mathrm{cont}})$  is autonomous and hence has output cardinality  $w_2$ . The hidden behavior  $\mathcal{N}_{\hat{w}_2}(\mathcal{O}_{\mathrm{cont}})$  is given by the latent variable representation

$$\begin{pmatrix} 0\\I \end{pmatrix} \hat{w}_2 = \begin{pmatrix} L_1(\frac{\mathrm{d}}{\mathrm{d}t})\\L_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix} l,$$

and hence, see e.g. [11, Lemma 2.9.5],

$$\operatorname{rank} \begin{pmatrix} 0 & L_1 \\ I & L_2 \end{pmatrix} - \operatorname{rank} \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} = \mathbf{w}_2.$$

On the other hand,

r

$$\operatorname{ank} \begin{pmatrix} 0 & L_1 \\ I & L_2 \end{pmatrix} = \mathbf{w}_2 + \operatorname{rank} \mathbf{L}_1$$

and hence

$$\operatorname{rank} \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} = \operatorname{rank} L_1.$$

The second statement now follows immediately.

For the special case of i/o-observers, the following theorem was previously announced in [16].

Theorem 5.3: Given an observer interconnection where the observer is nonintrusive, then  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  is autonomous if and only if  $\hat{w}_2$  is trackable from  $w_1$  (in  $\mathcal{O}$ ) and the observer behavior contains the controllable part of the plant behavior,  $\mathcal{P}_{\text{cont}} \subset \mathcal{O}$ .

*Proof:* Assume that  $\hat{w}_2$  is trackable from  $w_1$  (in  $\mathcal{O}$ ) and that  $\mathcal{P}_{\text{cont}} \subset \mathcal{O}$ . Consider the behavior  $(\mathcal{P}_{\text{aut}})_{w_1} \wedge_{w_1} \mathcal{O}$ , where  $(\mathcal{P}_{\text{aut}})_{w_1}$  is the projection of some autonomous part  $\mathcal{P}_{\text{aut}}$  of  $\mathcal{P}$  onto the variable  $w_1$ . By Lemma 5.1, the projection  $((\mathcal{P}_{\text{aut}})_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2}$  is autonomous. We will now prove that

$$\mathcal{E}(\mathcal{P}, \mathcal{O}) \subset ((\mathcal{P}_{\mathrm{aut}})_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2} + (\mathcal{P}_{\mathrm{aut}})_{w_2},$$

which implies that  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  is autonomous since the right hand side is an autonomous behavior. Indeed, let  $e \in \mathcal{E}(\mathcal{P}, \mathcal{O})$  and  $e = \hat{w}_2 - w_2$ . Then there exists  $w_1$  such that  $(w_1, w_2) \in \mathcal{P}$  and  $(w_1, \hat{w}_2) \in \mathcal{O}$ . Decompose  $(w_1, w_2) = (w_1^c, w_2^c) + (w_1^a, w_2^a)$ , where  $(w_1^c, w_2^c) \in \mathcal{P}_{cont}$  and  $(w_1^a, w_2^a) \in \mathcal{P}_{aut}$ . Since  $\mathcal{P}_{cont} \subset \mathcal{O}$  we have  $(w_1, \hat{w}_2) - (w_1^c, w_2^c) \in \mathcal{O}$ . The latter equals  $(w_1^a, \hat{w}_2 - w_2 + w_2^a) = (w_1^a, e + w_2^a)$ . Since this is in  $\mathcal{O}$ , we have  $e + w_2^a \in ((\mathcal{P}_{aut})_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2}$  and  $e \in ((\mathcal{P}_{aut})_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2} + (\mathcal{P}_{aut})_{w_2}$  as claimed.

Conversely, assume that  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  is autonomous. Then  $\hat{w}_2$  is trackable from  $w_1$  (in  $\mathcal{O}$ ), cf. Remark 4.8. Let

$$\mathcal{P}_{\mathrm{cont}} = \mathrm{im} \begin{pmatrix} M_1(\frac{\mathrm{d}}{\mathrm{d}t}) \\ M_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix}$$
 and  $\mathcal{O}_{\mathrm{cont}} = \mathrm{im} \begin{pmatrix} L_1(\frac{\mathrm{d}}{\mathrm{d}t}) \\ L_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix}$ 

be minimal (full column rank) image representations of the controllable parts of the plant behavior and of the observer behavior, respectively. Then the restricted error behavior  $\mathcal{E}(\mathcal{P}_{cont}, \mathcal{O}_{cont})$  is given by the latent variable representation

$$\begin{pmatrix} 0\\ I \end{pmatrix} e = \begin{pmatrix} M_1(\frac{\mathrm{d}}{\mathrm{d}t}) & -L_1(\frac{\mathrm{d}}{\mathrm{d}t}) \\ M_2(\frac{\mathrm{d}}{\mathrm{d}t}) & -L_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix} \begin{pmatrix} l\\ l' \end{pmatrix}.$$

Since  $\mathcal{P}_{\text{cont}} \subset \mathcal{P}$  and  $\mathcal{O}_{\text{cont}} \subset \mathcal{O}$ , it follows that  $(\mathcal{P}_{\text{cont}} \wedge_{w_1} \mathcal{O}_{w_1}) \wedge_{(w_2,\hat{w}_2)} \mathcal{D} \subset (\mathcal{P} \wedge_{w_1} \mathcal{O}) \wedge_{(w_2,\hat{w}_2)} \mathcal{D}$  and hence that  $\mathcal{E}(\mathcal{P}_{\text{cont}}, \mathcal{O}_{\text{cont}}) \subset \mathcal{E}(\mathcal{P}, \mathcal{O})$ . But then  $\mathcal{E}(\mathcal{P}_{\text{cont}}, \mathcal{O}_{\text{cont}})$  is

autonomous and has output cardinality  $w_2$ . Hence, see e.g. [11, Lemma 2.9.5],

$$\operatorname{rank} \begin{pmatrix} 0 & M_1 & -L_1 \\ I & M_2 & -L_2 \end{pmatrix} - \operatorname{rank} \begin{pmatrix} M_1 & -L_1 \\ M_2 & -L_2 \end{pmatrix} = \mathtt{w}_2.$$
(3)

By Lemma 5.2 there exist rational matrices S and X such that  $M_1 = L_1S$  and  $L_2 = XL_1$ . But then

$$\operatorname{rank} \begin{pmatrix} 0 & M_1 & -L_1 \\ I & M_2 & -L_2 \end{pmatrix} = \operatorname{rank} \begin{pmatrix} 0 & L_1S & -L_1 \\ I & M_2 & -L_2 \end{pmatrix}$$
$$= \operatorname{rank} \begin{pmatrix} 0 & 0 & -L_1 \\ I & M_2 - L_2S & -L_2 \end{pmatrix}$$
$$= \operatorname{rank} \begin{pmatrix} 0 & 0 & -L_1 \\ I & 0 & -L_2 \end{pmatrix}$$
$$= \operatorname{rank} \begin{pmatrix} 0 & 0 & -L_1 \\ I & 0 & -XL_1 \end{pmatrix}$$
$$= \operatorname{rank} \begin{pmatrix} 0 & 0 & -L_1 \\ I & 0 & -XL_1 \end{pmatrix}$$
$$= \operatorname{rank} \begin{pmatrix} 0 & 0 & -L_1 \\ I & 0 & 0 \end{pmatrix}$$
$$= \operatorname{w}_2 + \operatorname{rank} (-L_1).$$

Combining this with (3) yields

$$\operatorname{rank} \begin{pmatrix} M_1 & -L_1 \\ M_2 & -L_2 \end{pmatrix} = \operatorname{rank} (-L_1)$$
$$= \operatorname{rank} \begin{pmatrix} -L_1 \\ -XL_1 \end{pmatrix}$$
$$= \operatorname{rank} \begin{pmatrix} -L_1 \\ -L_2 \end{pmatrix}$$

and hence there exists a rational matrix T such that

$$\begin{pmatrix} M_1 \\ M_2 \end{pmatrix} = \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} T.$$

Factorize  $T = PQ^{-1}$  with P and Q polynomial. Obviously, the differential operator  $Q(\frac{d}{dt})$  is surjective. This implies that

$$\begin{aligned} \mathcal{P}_{\text{cont}} &= \operatorname{im} \begin{pmatrix} M_1(\frac{\mathrm{d}}{\mathrm{d}t}) \\ M_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix} Q(\frac{\mathrm{d}}{\mathrm{d}t}) \\ &= \operatorname{im} \begin{pmatrix} L_1(\frac{\mathrm{d}}{\mathrm{d}t}) \\ L_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix} P(\frac{\mathrm{d}}{\mathrm{d}t}) \subset \mathcal{O}_{\text{cont}} \subset \mathcal{O}. \end{aligned}$$

In analogy to similar results in geometric control, we refer to the previous result as an *internal model principle* for observers. Note that the condition in the theorem is necessary and sufficient and that we have not used nonintrusiveness in the proof of sufficiency. In the following we derive more refined versions of this principle for the cases of stable and trivial error behaviors, respectively.

We begin by refining Lemma 5.1.

Lemma 5.4: Let  $\mathcal{B}_1 \in \mathfrak{L}^{w_1}$  be a stable behavior and let  $\mathcal{B}_2 \in \mathfrak{L}^{w_1+w_2}$  be such that  $w_2$  is detectable from  $w_1$  (in  $\mathcal{B}_2$ ). Then  $(\mathcal{B}_1 \wedge_{w_1} \mathcal{B}_2)_{w_2}$  is stable.

*Proof:* Since  $\mathcal{B}_1$  is stable, it admits a kernel representation  $\mathcal{B}_1 = \ker(R(\frac{d}{dt}))$  where  $R(\lambda)$  has full column rank for all  $\lambda \in \mathbb{C}^+$ .  $w_2$  being detectable from  $w_1$  (in  $\mathcal{B}_2$ ) implies that  $\mathcal{N}_{w_2}(\mathcal{P})$  is stable and hence that  $\mathcal{B}_2$  admits a kernel representation  $\mathcal{B}_2 =$ 

 $\ker \left(R_1(\frac{\mathrm{d}}{\mathrm{d}t}) \quad R_2(\frac{\mathrm{d}}{\mathrm{d}t})\right)$  where  $R_2(\lambda)$  has full column rank for all  $\lambda \in \mathbb{C}^+$ . But then

$$\mathcal{B}_1 \wedge_{w_1} \mathcal{B}_2 = \ker \begin{pmatrix} R(\frac{\mathrm{d}}{\mathrm{d}t}) & 0\\ R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & R_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix}$$

and the matrix

$$\begin{pmatrix} R(\lambda) & 0\\ R_1(\lambda) & R_2(\lambda) \end{pmatrix}$$

has full column rank for all  $\lambda \in \mathbb{C}^+$ . Hence  $\mathcal{B}_1 \wedge_{w_1} \mathcal{B}_2$  is stable and so is its projection.

Next, we have a closer look at the special case of an antistable plant.

Proposition 5.5: Consider an observer interconnection where the plant is anti-stable and the observer is nonintrusive. If  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  is stable then  $\mathcal{P} \subset \mathcal{O}$ .

*Proof:* Let  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  be stable and let  $(w_1, w_2) \in \mathcal{P}$ . We need to prove that  $(w_1, w_2) \in \mathcal{O}$ . Since  $\mathcal{O}$  is nonintrusive there exists  $\hat{w}_2$  such that  $(w_1, \hat{w}_2) \in \mathcal{O}$ . It follows that  $e = \hat{w}_2 - w_2$  is a stable Bohl function. Since  $\mathcal{P}$  is anti-stable its nonzero trajectories are anti-stable Bohl functions. Hence either  $w_2 = 0$  or  $w_2$  is an anti-stable Bohl function and similarly for  $w_1$ .

Assume  $w_2 = 0$ , then  $\hat{w}_2 = e$  is a stable Bohl function. Let  $\mathcal{O} = \ker \left(\hat{R}_1(\frac{d}{dt}) \ \hat{R}_2(\frac{d}{dt})\right)$  be a kernel representation. Then  $\hat{R}_1(\frac{d}{dt})w_1 = -\hat{R}_2(\frac{d}{dt})\hat{w}_2$  where the left hand side is either equal to zero or an anti-stable Bohl function and the right hand side is a stable Bohl function. It follows that  $\hat{R}_1(\frac{d}{dt})w_1 = 0$  and hence that  $(w_1, w_2) = (w_1, 0) \in \mathcal{O}$  in this case  $(w_2 = 0)$ .

We just proved that  $\mathcal{N}_{w_1}(\mathcal{P}) \subset \mathcal{N}_{w_1}(\mathcal{O})$ . Let  $\mathcal{P} = \ker \left(R_1(\frac{\mathrm{d}}{\mathrm{d}t}) \mid R_2(\frac{\mathrm{d}}{\mathrm{d}t})\right)$  be a kernel representation, then there exists a polynomial matrix S such that  $\hat{R}_1 = SR_1$ .

Assume now that  $w_2$  is an anti-stable Bohl function (the alternative case). Using the above kernel representations it follows that  $\hat{R}_2(\frac{d}{dt})e = \hat{R}_2(\frac{d}{dt})\hat{w}_2 - \hat{R}_2(\frac{d}{dt})w_2 =$  $-S(\frac{d}{dt})R_1(\frac{d}{dt})w_1 - \hat{R}_2(\frac{d}{dt})w_2 = (S(\frac{d}{dt})R_2(\frac{d}{dt}) - \hat{R}_2(\frac{d}{dt}))w_2$ where the left hand side is a stable Bohl function and the right hand side is either equal to zero or an anti-stable Bohl function. It follows that  $\hat{R}_2(\frac{d}{dt})(\hat{w}_2 - w_2) = 0$  and hence that  $(0, \hat{w}_2 - w_2) \in \mathcal{O}$ . But this implies that  $(w_1, w_2) =$  $(w_1, \hat{w}_2) - (0, \hat{w}_2 - w_2) \in \mathcal{O}$ . This concludes the proof.

Theorem 5.6: Given an observer interconnection where the observer is nonintrusive, then  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  is stable if and only if  $\hat{w}_2$  is detectable from  $w_1$  (in  $\mathcal{O}$ ) and the observer behavior contains the anti-stabilizable part of the plant behavior,  $\mathcal{P}_{\text{antistab}} \subset \mathcal{O}$ .

*Proof:* Choose a a controllable/autonomous decomposition  $\mathcal{P} = \mathcal{P}_{cont} \oplus \mathcal{P}_{aut}$  and the associated anti-stable/stable decomposition  $\mathcal{P}_{aut} = \mathcal{P}_+ \oplus \mathcal{P}_-$ , cf. Proposition 2.2.

Assume now that  $\hat{w}_2$  is detectable from  $w_1$  (in  $\mathcal{O}$ ) and that  $\mathcal{P}_{\text{antistab}} \subset \mathcal{O}$ . Consider the behavior  $(\mathcal{P}_-)_{w_1} \wedge_{w_1} \mathcal{O}$ . By Lemma 5.4, the projection  $((\mathcal{P}_-)_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2}$  is stable. We will now prove that

$$\mathcal{E}(\mathcal{P},\mathcal{O}) \subset ((\mathcal{P}_{-})_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2} + (\mathcal{P}_{-})_{w_2}$$

which implies that  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  is stable since the right hand side is a stable behavior. Indeed, let  $e \in \mathcal{E}(\mathcal{P}, \mathcal{O})$  and  $e = \hat{w}_2 - w_2$ . Then there exists  $w_1$  such that  $(w_1, w_2) \in \mathcal{P}$  and  $(w_1, \hat{w}_2) \in \mathcal{O}$ . Decompose  $(w_1, w_2) = (w'_1, w'_2) + (w''_1, w''_2)$  where  $(w'_1, w'_2) \in \mathcal{P}_{cont} \oplus \mathcal{P}_+$  and  $(w''_1, w''_2) \in \mathcal{P}_-$ . By Proposition 2.4,  $\mathcal{P}_{antistab} = \mathcal{P}_{cont} \oplus \mathcal{P}_+$ . Since  $\mathcal{P}_{antistab} \subset \mathcal{O}$ we have  $(w_1, \hat{w}_2) - (w'_1, w'_2) \in \mathcal{O}$ . The latter equals  $(w''_1, \hat{w}_2 - w_2 + w''_2) = (w''_1, e + w''_2)$ . Since this is in  $\mathcal{O}$ , we have  $e + w''_2 \in ((\mathcal{P}_-)_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2}$  and  $e \in ((\mathcal{P}_-)_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2} + (\mathcal{P}_-)_{w_2}$  as claimed.

Conversely, assume that  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  is stable. Then  $\hat{w}_2$  is detectable from  $w_1$  (in  $\mathcal{O}$ ), cf. Remark 4.8. Since  $\mathcal{E}(\mathcal{P}, \mathcal{O})$  is autonomous it follows from Theorem 5.3 that  $\mathcal{P}_{cont} \subset \mathcal{O}$ . Since  $\mathcal{P}_+ \subset \mathcal{P}, \mathcal{E}(\mathcal{P}, \mathcal{O})$  stable implies that  $\mathcal{E}(\mathcal{P}_+, \mathcal{O})$  is stable. Moreover,  $\mathcal{O}$  is clearly also nonintrusive with respect to  $\mathcal{P}_+$ . It follows from Proposition 5.5 that  $\mathcal{P}_+ \subset \mathcal{O}$  and hence by Proposition 2.4 that  $\mathcal{P}_{antistab} = \mathcal{P}_{cont} \oplus \mathcal{P}_+ \subset \mathcal{O}$ .

Again, we have not used nonintrusiveness in the proof of sufficiency. We finally turn to the case of exact observers where we obtain the following "full" internal model principle.

Theorem 5.7: Given an observer interconnection where the observer is nonintrusive, then  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \{0\}$  if and only if  $\hat{w}_2$  is observable from  $w_1$  (in  $\mathcal{O}$ ) and the observer behavior contains the plant behavior,  $\mathcal{P} \subset \mathcal{O}$ .

*Proof:* Assume that  $\hat{w}_2$  is observable from  $w_1$  (in  $\mathcal{O}$ ) and that  $\mathcal{P} \subset \mathcal{O}$ . Let  $\mathcal{P} = \ker \left( R_1(\frac{d}{dt}) \quad R_2(\frac{d}{dt}) \right)$  be a minimal kernel representation and let  $\mathcal{O} = \ker \left( \hat{R}_1(\frac{d}{dt}) \quad \hat{R}_2(\frac{d}{dt}) \right)$ be any kernel representation. Then there exists a polynomial matrix S such that  $\left( \hat{R}_1 \quad \hat{R}_2 \right) = \left( SR_1 \quad SR_2 \right)$ . Furthermore,  $\mathcal{N}_{\hat{w}_2}(\mathcal{O}) = \{0\}$  implies that  $S(\lambda)R_2(\lambda)$  has full column rank for all  $\lambda \in \mathbb{C}$ . By Lemma 3.4,  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \ker(S(\frac{d}{dt})R_2(\frac{d}{dt}))$ and hence  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \{0\}$ .

Conversely, assume that  $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \{0\}$ . Then  $\hat{w}_2$  is observable from  $w_1$  (in  $\mathcal{O}$ ), cf. Remark 4.8. Let  $(w_1, w_2) \in \mathcal{P}$ then there exists  $\hat{w}_2$  such that  $(w_1, \hat{w}_2) \in \mathcal{O}$ . But then  $\hat{w}_2 - w_2 = e \in \mathcal{E}(\mathcal{P}, \mathcal{O}) = \{0\}$  and hence  $\hat{w}_2 = w_2$ . It follows that  $(w_1, w_2) = (w_1, \hat{w}_2) \in \mathcal{O}$  and hence that  $\mathcal{P} \subset \mathcal{O}$ .

Note that the last theorem implies that exact observers are necessarily consistent, an observation already made in [3]. The common theme of the previous three theorems could be summed up as follows. Given an observer interconnection where the observer is nonintrusive, the observer behavior must necessarily contain that part of the plant behavior that we do not want to be present in the associated error behavior.

*Remark 5.8:* The internal model principle can be used to derive parametrizations of all nonintrusive tracking (stable, exact) observers for a given plant, cf. also [2], [12]. We only sketch the general idea, and only for the tracking case. The interested reader can find a full account of the resulting parametrizations in [17]. Consider a minimal kernel representation  $\mathcal{P} = \ker \left( R(\frac{d}{dt}) \right)$  and let U and V be unimodular matrices such that  $URV = (D \ 0)$ , the Smith form of R. Then the controllable subbehavior is given by  $\mathcal{P}_{cont} = \ker \left( (I \ 0) V(\frac{d}{dt})^{-1} \right)$ . By Theorem 5.3, all nonintrusive tracking observers are of the form

$$\mathcal{O} = \ker \left( S \begin{pmatrix} I & 0 \end{pmatrix} V \begin{pmatrix} \frac{\mathrm{d}}{\mathrm{d}t} \end{pmatrix}^{-1} \right),$$

where S is a polynomial matrix such that  $\hat{w}_2$  is trackable from  $w_1$  (in  $\mathcal{O}$ ). The latter condition can be formulated in terms of a column rank condition, although the details require some cumbersome notation. This is because the block decomposition

of the observer variables need not be compatible with the block decomposition in the Smith form above.

#### VI. UNKNOWN INPUT OBSERVERS

In this section we provide an application of the behavioral theory to unknown input observer design for descriptor systems. More specifically, we show that the (full-order) "proportional-integral" observer derived by Koenig and Mammar [18] is an example of an observer that contains a full internal model. More general observer designs could be achieved using the more specific internal model requirement from Theorem 5.6.

Koenig and Mammar base their observer design on the following model, see Equation (4) in [18], that is assumed to be impulse observable and R-detectable (cf. Definition 1 in [18]).

$$\begin{bmatrix} E & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{f} \end{bmatrix} = \begin{bmatrix} A & N \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ f \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u,$$

$$y = \begin{bmatrix} C & 0 \end{bmatrix} \begin{bmatrix} x \\ f \end{bmatrix},$$
(4)

where E, A, N, B and C are constant matrices, x is the to be observed state, f is a to be observed unknown input (fault), u is a measured input and y is a measured output. The behavior of (4) is  $\mathcal{P} = \ker \left(R_1\left(\frac{d}{dt}\right) \quad R_2\left(\frac{d}{dt}\right)\right)$  with variables  $(w_1, w_2) = ((y, u), (x, f))$  and

$$R_1(\frac{\mathrm{d}}{\mathrm{d}t}) = \begin{pmatrix} 0 & -B\\ 0 & 0\\ -I & 0 \end{pmatrix}, \quad R_2(\frac{\mathrm{d}}{\mathrm{d}t}) = \begin{pmatrix} \frac{\mathrm{d}}{\mathrm{d}t}E - A & -N\\ 0 & \frac{\mathrm{d}}{\mathrm{d}t}I\\ C & 0 \end{pmatrix}.$$

The observer provided in Equation (5) of [18] takes the form

/ \

$$\begin{pmatrix} -T_2 & 0 & I & 0 & -M_1 \\ -L_3 & 0 & L_3C & \frac{d}{dt}I & 0 \\ L_1 + L_2 & J & 0 & T_1N & -\frac{d}{dt}I + F \end{pmatrix} \begin{pmatrix} y \\ u \\ \hat{x} \\ \hat{f} \\ z \end{pmatrix} = 0,$$
(5)

where  $\hat{x}$  and  $\hat{f}$  are estimates for x and f, respectively, and z is an auxiliary variable. According to Algorithm 1 in [18], the observer matrices are chosen such that  $\begin{bmatrix} T_1 & T_2 \end{bmatrix} \begin{bmatrix} E^\top & C^\top \end{bmatrix}^\top = I$ ,  $F = T_1A - L_2C$ ,  $L_1 = FT_2$ ,  $J = T_1B$ ,  $M_1 = I$  and

$$\begin{bmatrix} T_1 A & T_1 N \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} L_2 \\ L_3 \end{bmatrix} \begin{bmatrix} C & 0 \end{bmatrix} = \begin{bmatrix} F & T_1 N \\ -L_3 C & 0 \end{bmatrix}$$
 is stable. (6)

This choice is possible due to the assumptions of impulse observability and R-detectability. The unimodular transformation

$$\begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\frac{d}{dt}I + F & 0 & I \end{pmatrix} \begin{pmatrix} -T_2 & 0 & I & 0 & -M_1 \\ -L_3 & 0 & L_3C & \frac{d}{dt}I & 0 \\ L_1 + L_2 & J & 0 & T_1N & -\frac{d}{dt}I + F \end{pmatrix}$$
$$= \begin{pmatrix} -T_2 & 0 & I & 0 & -M_1 \\ -L_3 & 0 & L_3C & \frac{d}{dt}I & 0 \\ \frac{d}{dt}T_2 - FT_2 + L_1 + L_2 & J & -\frac{d}{dt}I + F & T_1N & 0 \end{pmatrix}$$

allows to eliminate the variable z from the observer equation to obtain the observer behavior  $\mathcal{O} = \ker \left( \hat{R}_1(\frac{d}{dt}) \quad \hat{R}_2(\frac{d}{dt}) \right)$  with variables  $(w_1, \hat{w}_2) = ((y, u), (\hat{x}, \hat{f}))$  and

$$\hat{R}_1(\frac{\mathrm{d}}{\mathrm{d}t}) = \begin{pmatrix} -L_3 & 0\\ \frac{\mathrm{d}}{\mathrm{d}t}T_2 + L_2 & T_1B \end{pmatrix},$$
$$\hat{R}_2(\frac{\mathrm{d}}{\mathrm{d}t}) = \begin{pmatrix} L_3C & \frac{\mathrm{d}}{\mathrm{d}t}I\\ -\frac{\mathrm{d}}{\mathrm{d}t}I + F & T_1N \end{pmatrix},$$

where we have used the defining equations for  $L_1$ , J and  $M_1$ . By inspection,

$$\left( S(\frac{\mathrm{d}}{\mathrm{d}t}) R_1(\frac{\mathrm{d}}{\mathrm{d}t}) \quad S(\frac{\mathrm{d}}{\mathrm{d}t}) R_2(\frac{\mathrm{d}}{\mathrm{d}t}) \right) = \left( \hat{R}_1(\frac{\mathrm{d}}{\mathrm{d}t}) \quad \hat{R}_2(\frac{\mathrm{d}}{\mathrm{d}t}) \right)$$

for the choice

$$S(\frac{\mathrm{d}}{\mathrm{d}t}) = \begin{pmatrix} 0 & I & L_3\\ -T_1 & 0 & -\frac{\mathrm{d}}{\mathrm{d}t}T_2 - L_2 \end{pmatrix}$$

and hence  $\mathcal{P} \subset \mathcal{O}$ . Here we have used the defining equations for  $\begin{pmatrix} T_1 & T_2 \end{pmatrix}$  and F. It is instructive to see how the various choices in Algorithm 1 of [18] correspond to the (full) internal model property of the observer behavior. The observer (5) is now simply an input/state/output realization of the observer behavior  $\mathcal{O}$  which is hence nonintrusive. Detectability of  $\hat{w}_2$ from  $w_1$  (in  $\mathcal{O}$ ) follows from condition (6), and Theorem 5.6 applies.

## VII. THE STATE SPACE CASE

In this section we provide a link from our results to classical results from state observer theory and discuss the relationship between partial state observers and simulation relations. Note that the results in Sections III-V are *completely independent* of the types of variables involved, whether they are states, inputs, or else, as well as of the specific behavior representation in form of equations. This is demonstrated in the example in Section VI. This present section serves to illustrate how these results can be specialized and applied to a particular example, namely partial state observation from inputs and outputs where the observed system is given in input/state/output form, and where we are seeking a representation of the observer from the same class. We chose this particular example since most readers would be familiar with this setup and refer to Section VI and [3] for other examples, namely an application of the behavioral theory to unknown input observers and to fault detection and isolation.

Consider a plant whose full behavior  $\mathcal{P}_{\text{full}}$  with variables  $(u, x, y, z) \in \mathbb{R}^{m+n+p+k}$  is given by

$$\dot{x} = Ax + Bu, 
y = Cx, 
z = Vx.$$
(7)

Here, the various matrices are constant matrices. We denote the projection onto the variables (u, y, z) by  $\mathcal{P} = (\mathcal{P}_{\text{full}})_{(u,y,z)}$ . Consider candidate observers for z from (u, y) whose full behavior  $\mathcal{O}_{\text{full}}$  with variables  $(u, y, v, \hat{z}) \in \mathbb{R}^{m+p+q+k}$  is given by

$$\dot{v} = Kv + Ly + Mu,$$
  
$$\dot{z} = Pv + Qy.$$
 (8)

Again, the various matrices are all constant and we assume that the matrix pair (P, K) is observable. We denote the projection onto the variables  $(u, y, \hat{z})$  by  $\mathcal{O} = (\mathcal{O}_{\text{full}})_{(u, y, \hat{z})}$ . We are interested in the observer *characterization* problem, i.e. in an answer to the question when a *given* observer of the form (8) asymptotically estimates z. Note that in this problem we can not change the observer equation without changing the problem. This is different to the observer *existence* problem where we are only interested whether or not we can build *some* observer of the form (8). For the latter problem, it is common practice to make a series of assumptions without loss of generality, see e.g. [19]. See the discussion in [20, Section 4] for the exact relationship between the two problems. In particular, [20, Example 12] is an example for an observer where the usual assumptions (such as P = I) are *not* without loss of generality for the observer *characterization* problem.

Proposition 7.1:  $\mathcal{P} = (\mathcal{P})_{(u,y,z)} \subset (\mathcal{O})_{(u,y,\hat{z})} = \mathcal{O}$  if and only if there exists a (constant) matrix U such that

$$UA - KU = LC,$$
  

$$M = UB,$$
  

$$V = PU + QC.$$
(9)

In this case, the error dynamics are given by

where d = v - Ux and  $e = \hat{z} - z$ .

*Proof:* Similar to [12, Theorem 5.1], equation (9) can be rewritten into the form

$$\begin{pmatrix} U & 0 & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} \frac{\mathrm{d}}{\mathrm{d}t}I - A & -B & 0 & 0 \\ C & 0 & -I & 0 \\ V & 0 & 0 & -I \end{pmatrix} = \\ \begin{pmatrix} \frac{\mathrm{d}}{\mathrm{d}t}I - K & -M & -L & 0 \\ P & 0 & Q & -I \end{pmatrix} \begin{pmatrix} U & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ C & 0 & 0 & 0 \\ 0 & 0 & 0 & I \end{pmatrix},$$

implying that the map

$$i: \mathcal{P}_{\mathrm{full}} \longrightarrow \mathcal{O}_{\mathrm{full}}, \ \begin{pmatrix} x \\ u \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} v \\ u \\ y \\ \hat{z} \end{pmatrix} = \begin{pmatrix} Ux \\ u \\ y \\ z \end{pmatrix}$$

is a (continuous) behavior homomorphism. This map restricts to the inclusion  $\mathcal{P} \subset \mathcal{O}$ . Conversely,  $\mathcal{P} \subset \mathcal{O}$  implies Equation (9) by a straightforward modification of Theorem 3.4 and Theorem 3.9 in [21]. Equation (10) follows from a simple direct calculation using (7), (8) and (9).

A simpler version of the Sylvester-type Equation (9) can already be found in Luenberger's original paper [1, Theorem 1]. In the terminology of geometric control, it implies that ker(U)is a conditioned invariant subspace contained in ker(V) and with outer spectrum equal to the spectrum of K (see, e.g., [22]).

We can now apply our internal model principle to obtain the following characterization of asymptotic observers.

Theorem 7.2: Let all uncontrollable modes in (7) be unstable and let (P, K) in (8) be observable. Then (8) is an asymptotic observer for z from (u, y) if and only if K is

Hurwitz and there exists a (constant) matrix U such that (9) holds.

**Proof:** The observer  $\mathcal{O} = (\mathcal{O}_{\text{full}})_{(u,y,\hat{z})}$  is an i/o-observer and hence nonintrusive. Since all uncontrollable modes in (7) are unstable, any autonomous complement of the controllable part of  $\mathcal{P} = (\mathcal{P}_{\text{full}})_{(u,y,z)}$  is automatically anti-stable. Hence the anti-stabilizable part  $\mathcal{P}_{\text{antistab}} = \mathcal{P}$ . Since (P, K) is observable,  $\hat{z}$  is detectable from (u, y) (in  $\mathcal{O} = (\mathcal{O}_{\text{full}})_{(u,y,\hat{z})}$ ) if and only if K is Hurwitz. The result now follows from Theorem 5.6 and Proposition 7.1.

Theorem 7.2 is sometimes assumed to be a classical result, and indeed, sufficiency was already established in Luenberger's work, starting with [1]. That work, however, does not study necessity, except for observers that additionally have the tracking property, leaving the difficult part of the question whether every asymptotic observer necessarily contains an internal model unresolved. Subsequent notable necessity results in the literature are Fortmann and Williamson [23] where a stronger notion of asymptotic observers is used (asymptotically matching derivatives of all orders) and Moore and Ledwich [24] whose proof uses a misshaped "reachability matrix", leading to a partially wrong result. Several books have appeared that contain quotes of those results without improving on the proofs, e.g. [25], [26]. Many subsequent papers cite these books but often contain only statements for the special case P = I, e.g. [27], [19]. The first full and correct proof of the necessity part the authors know of is relatively recent [28] and only treats the case of a controllable observed system. See [20, Section 3.1] for a more in-depth discussion of the classical literature on this topic.

It follows from Theorem 7.2 that partial state observers of the above form *necessarily* simulate the plant dynamics in the following precise sense. We start the discussion by recalling the usual notion of simulation relation between a pair of state space systems (see, e.g., [29]).

Definition 7.3: Given two systems  $\Sigma_i$ , i = 1, 2 of the form

$$\begin{aligned} \dot{x}_i &= A_i x_i + B_i u_i, \quad x_i(t) \in \mathcal{X}_i, u_i(t) \in \mathcal{U} \text{ for all } t, \\ y_i &= C_i x_i, \quad y_i(t) \in \mathcal{Y} \text{ for all } t, \end{aligned}$$

a simulation relation of  $\Sigma_1$  by  $\Sigma_2$  is a linear subspace  $S \subset \mathcal{X}_1 \times \mathcal{X}_2$  with the property that for all pairs of initial values  $(x_1(0), x_2(0)) \in S$  and all joint input functions  $u_1(\cdot) = u_2(\cdot)$  the resulting trajectories satisfy  $(x_1(t), x_2(t)) \in S$  and  $y_1(t) = y_2(t)$  for all  $t \ge 0$ .

In words,  $\Sigma_2$  simulates  $\Sigma_1$  with respect to the simulation relation S if a related pair of initial values and a joint input gives rise to related state trajectories and the same output. Adapting this definition in the obvious way we can define simulation relations for plant/observer pairs.

Definition 7.4: A simulation relation of the plant (7) by the observer (8) is a linear subspace  $S \subset \mathbb{R}^n \times \mathbb{R}^q$  with the property that for all pairs of initial values  $(x(0), v(0)) \in S$  and all input functions  $u(\cdot)$  the resulting trajectories satisfy  $(x(t), v(t)) \in S$  and  $z(t) = \hat{z}(t)$  for all  $t \ge 0$ .

Note that in this situation the common input u and the output/input y are joint through the observer interconnection structure. This situation is hence not captured by the classical

definition of simulation relations but by a behavioral generalization of the same, see [30]. The following is now a direct consequence of Theorem 7.2 and Proposition 7.1.

Corollary 7.5: Let all uncontrollable modes in (7) be unstable and let (8) be an observable asymptotic observer for z from (u, y). Then the observer (8) simulates the plant (7) with respect to the simulation relation  $S = \ker(-U I)$  where U is as in (9).

This corollary provides a state space interpretation of the internal model principle for observers.

#### VIII. CONCLUSIONS

In this paper we have studied the observer problem in a behavioral framework. For a given plant behavior, with a partition of the system variable into a set of measured components and a set of to be estimated components, we have characterized all error behaviors that can be achieved in an observer interconnection. Using this characterization we have re-established the necessary and sufficient conditions for the existence of tracking, stable and exact observers. As main results of this paper, we have established behavioral formulations of an internal model principle for observers. For nonintrusive observers we have shown that the error behavior is autonomous if and only if in the observer the estimator variable is trackable from the measured one, and the observer behavior contains the controllable part of the plant. An observer is asymptotic if and only if the estimator variable is detectable from the measured one, and the observer contains the anti-stabilizable part of the plant. Finally, the observer is exact if and only if the estimator variable is observable from the measured one, and the observer contains the entire plant behavior. We have applied our results to unknown input observer design for descriptor systems and to the case where the plant is strictly proper and represented in input-state-output form. Finally, in the state space context, we have given an interpretation of the internal model principle in terms of simulation of the plant by the observer.

### IX. ACKNOWLEDGMENTS

The authors would like to thank an anonymous reviewer for pointing out an embarrassing omission in one of the main proofs in a previous version of this paper.

This work was partially supported by the Australian Research Council through the ARC Discovery Project DP120100316 "Geometric observer theory for mechanical control systems".

The SISTA research program of the K.U. Leuven is supported by the Research Council KUL under projects: GOA AMBioR-ICS, CoE EF/05/006 Optimization in Engineering (OPTEC), IOF-SCORES4CHEM; by the Flemish Government: FWO: 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), G.0320.08 (convex MPC), G.0558.08 (Robust MHE), G.0557.08 (Glycemia2), G.0588.09 (Brain-machine) research communities (IC-CoS, ANMMM, MLDM); G.0377.09 (Mechatronics MPC) and by IWT: McKnow-E, Eureka-Flite+, SBO LeCoPro, SBO Climaqs; by the Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011) ; by the EU: ERNSI; FP7-HD-MPC (INFSO- ICT-223854); and by several contract research projects.

#### REFERENCES

- [1] D. G. Luenberger, "Observing the state of a linear system," *IEEE Transactions on Military Electronics*, vol. 8, no. 2, pp. 74–80, 1964.
- [2] M. E. Valcher and J. C. Willems, "Observer synthesis in the behavioral approach," *IEEE Transactions on Automatic Control*, vol. 44, no. 12, pp. 2297–2307, 1999.
- [3] M. Bisiacco, M. E. Valcher, and J. C. Willems, "A behavioral approach to estimation and dead-beat observer design with applications to state space models," *IEEE Transactions on Automatic Control*, vol. 51, no. 11, pp. 1787–1797, 2006.
- [4] J. C. Willems, "The behavioral approach to open and interconnected systems," *IEEE Control Systems Magazine*, vol. 27, no. 6, pp. 46–99, 2007.
- [5] M. N. Belur and H. L. Trentelman, "Stabilization, pole placement, and regular implementability," *IEEE Transactions on Automatic Control*, vol. 47, no. 5, pp. 735–744, 2002.
- [6] A. J. van der Schaft, "Achievable behavior of general systems," Systems and Control Letters, vol. 49, no. 2, pp. 141–149, 2003.
- [7] J. Trumpf, H. L. Trentelman, and J. C. Willems, "An internal model principle for observers," in *Proceedings of the 50th IEEE Conference* on Decision and Control, 2011, pp. 3992–3999.
- [8] V. Lomadze, "Axiomatic characterization of linear differential systems (and operators)," *Automatica*, vol. 48, no. 5, pp. 815–819, 2012.
- [9] J. W. Polderman and J. C. Willems, *Introduction to mathematical systems theory: a behavioral approach*. Springer, 1998.
- [10] S. V. Gottimukkala, S. Fiaz, and H. L. Trentelman, "Equivalence of rational representations of behaviors," *Systems and Control Letters*, vol. 60, no. 2, pp. 119–127, 2011.
- [11] M. N. Belur, "Control in a behavioral context," Ph.D. dissertation, Rijksuniversiteit Groningen, 2003.
- [12] P. A. Fuhrmann, "Observer theory," *Linear Algebra and its Applications*, vol. 428, no. 1, pp. 44–136, 2008.
- [13] M. L. J. Hautus, "Strong detectability and observers," *Linear Algebra and its Applications*, vol. 50, pp. 353–368, 1983.
- [14] I. Blumthaler and U. Oberst, "T-observers," *Linear Algebra and its Applications*, vol. 430, no. 8-9, pp. 2416–2447, 2009.
- [15] I. Blumthaler, "Functional T-observers," *Linear Algebra and its Applications*, vol. 432, no. 6, pp. 1560–1577, 2010.
- [16] J. Trumpf and J. C. Willems, "An internal model principle for observers," Book of Abstracts, Characteristic Functions and Transfer Functions in Operator Theory and System Theory: a conference dedicated to Paul Fuhrmann on his 70th anniversary and to the memory of Moshe Livsic on his 90th anniversary, July 2007.
- [17] I. Blumthaler and J. Trumpf, "A new parametrization of observers," in *Proceedings of the 20th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, 2012, paper no. 99 (no pagination).
- [18] D. Koenig and S. Mammar, "Design of proportional-integral observer for unknown input descriptor systems," *IEEE Transactions on Automatic Control*, vol. 47, no. 12, pp. 2057–2062, 2002.
- [19] T. L. Fernando, L. S. Jennings, and H. M. Trinh, "Generality of functional observer structures," in *Proceedings of the 50th IEEE Conference* on Decision and Control, 2011, pp. 4000–4004.
- [20] J. Trumpf, "On state observers," in Mathematical System Theory Festschrift in Honor of Uwe Helmke on the Occasion of his Sixtieth Birthday, K. Hüper and J. Trumpf, Eds. CreateSpace, to appear 2013, available online http://users.cecs.anu.edu.au/~trumpf/UH60Festschrift. pdf.
- [21] —, "Observers for linear time-varying systems," *Linear Algebra and its Applications*, vol. 425, no. 2–3, pp. 303–312, 2007.
- [22] —, "On the geometry and parametrization of almost invariant subspaces and observer theory," Ph.D. dissertation, Universität Würzburg, 2002.
- [23] T. E. Fortmann and D. Williamson, "Design of low-order observers for linear feedback control laws," *IEEE Transactions on Automatic Control*, vol. 17, no. 3, pp. 301–308, 1972.
- [24] J. B. Moore and G. F. Ledwich, "Minimal order observers for estimating linear functions of a state vector," *IEEE Transactions on Automatic Control*, vol. 20, no. 5, pp. 623–632, 1975.
- [25] J. O'Reilly, Observers for linear systems. Academic Press, 1983.
- [26] C.-T. Chen, *Linear system theory and design*. Harcourt Brace College Publishers, 1984.
- [27] M. Darouach, "Existence and design of functional observers for linear systems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 940–943, 2000.

- [28] P. A. Fuhrmann and U. Helmke, "On the parametrization of conditioned invariant subspaces and observer theory," *Linear Algebra and its Applications*, vol. 332–334, pp. 265–353, 2001.
- [29] A. J. van der Schaft, "Equivalence of dynamical systems by bisimulation," *IEEE Transactions on Automatic Control*, vol. 49, no. 12, pp. 2160–2172, 2004.
- [30] A. A. Julius and A. J. van der Schaft, "Bisimulation as congruence in the behavioral setting," in *Proceedings of the 44th IEEE Conference on Decision and Control*, 2005, pp. 814–819.



**Jochen Trumpf** (M'04) received the Dipl.-Math. and Dr. rer. nat. degrees in mathematics from the University of Würzburg, Germany, in 1997 and 2002, respectively. He is currently a Fellow in the Research School of Engineering at the Australian National University. His research interests include observer theory and design, linear systems theory and optimisation on manifolds with applications in robotics, computer vision and wireless communication.



Harry L. Trentelman (SM'98) is a full professor in Systems and Control at the Johann Bernoulli Institute for Mathematics and Computer Science of the University of Groningen in The Netherlands. From 1991 to 2008, he served as an associate professor and later as an adjoint professor at the same institute. From 1985 to 1991, he was an assistant professor, and later, an associate professor at the Mathematics Department of the University of Technology at Eindhoven, The Netherlands. He obtained his Ph.D. degree in Mathematics at the University

of Groningen in 1985. His research interests are the behavioral approach to systems and control, robust control, model reduction, multi-dimensional linear systems, hybrid systems, analysis, control and model reduction of networked systems, and the geometric theory of linear systems. He is a co-author of the textbook Control Theory for Linear Systems (Springer, 2001). Dr. Trentelman is an associate editor of the IEEE Transactions on Automatic Control and of Automatica, and is past associate editor of the SIAM Journal on Control and Optimization and of Systems and Control Letters.



Jan C. Willems (S'66 - M'68 - SM'78 - F'80 - LF'05) was born in Bruges, Flanders, Belgium. He graduated from the University of Gent in 1963 in electro-mechanical engineering. He obtained an M.Sc. degree from the University of Rhode Island and the Ph.D. degree from MIT. He was on the electrical engineering faculty of MIT from 1968 to 1973, including a one-year postdoctoral stay at the Department of Applied Mathematics and Theoretical Physics of Cambridge University. From 1973 until 2003 he was professor of Systems and Control

with the Mathematics Department at the University of Groningen in The Netherlands. In 2003, he became emeritus professor and was subsequently a Guest Professor with the Department of Electrical Engineering at the K.U. Leuven. More details on his CV, a description of his research activities, and a list of publications can be found on his website http://homes.esat.kuleuven.be/~jwillems.