An internal model principle for observers

Trumpf, Jochen; Trentelman, Harry L.; Willems, Jan C.

Published in:
Proceedings of the 50th IEEE Conference on Decision and Control

2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

An internal model principle 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 controllable part and the antistable part of some autonomous 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 explicitely 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 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 the state space context. In [3], additional results were obtained in the context of discrete-time behaviors, specifically on the existence of deadbeat observers.

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 [4], [5], 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 controllable part and the antistable part of some autonomous part of the plant behavior. We also formulate refinements of this internal model principle for tracking observers and for exact observers.

We will briefly sketch how to use the internal model principle to obtain parametrizations of all *nonintrusive* tracking (stable, exact) observers for a given plant. Because of space limitations, this parametrization will only be sketched for tracking observers. We will conclude this paper with an application of the behavioral results to the case of strictly proper input-output systems represented in state space form.

To conclude this section, some words on notation and

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

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

J.C. Willems is with the Department of Electrical Engineering, Katholieke Universiteit Leuven, B-3001, Leuven-Heverlee, Belgium. Email: Jan.Willems@esat.kuleuven.be

978-1-61284-801-3/11/$26.00 ©2011 IEEE

nomenclature used. We use the standard symbols for the fields of real and complex numbers $\mathbb{R}$ and $\mathbb{C}$. We use $\mathbb{R}^{\mathtt{n}}$, $\mathbb{R}^{\mathtt{n}\times\mathtt{m}}$, etc. for the real linear spaces of vectors and matrices with components in $\mathbb{R}$. Often, the notation $\mathbb{R}^{\mathtt{w}}$, $\mathbb{R}^{\mathtt{w}_1}$, ... is used if w, w_1, ... denote typical elements of that vector space, or typical functions taking their values in that vector space. $\mathfrak{C}^\infty(\mathbb{R},\mathbb{R}^{\mathtt{w}})$ will denote the set of infinitely often differentiable functions from $\mathbb{R}$ to $\mathbb{R}^{\mathtt{w}}$. $\mathbb{R}[\xi]$ denotes the ring of polynomials in the indeterminate ξ with real coefficients. We use $\mathbb{R}^{\mathtt{n}\times\mathtt{m}}[\xi]$ for the space of matrices with components in $\mathbb{R}[\xi]$. Elements of $\mathbb{R}^{\mathtt{n}\times\mathtt{m}}[\xi]$ are called *real polynomial matrices*.

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.

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 Σ_1 and Σ_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 Σ_1 and Σ_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* Σ_1 *and* Σ_2 *through* c as the dynamical system

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

with interconnected behavior

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

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 denote the interconnected behavior $\mathfrak{B}$ by $\mathfrak{B}_1 \wedge_c \mathfrak{B}_2$.

In this paper we will restrict ourselves to linear time-invariant 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, $\mathbb{R}^{\mathtt{w}}$; correspondingly, the manifest variable is then given as $w = \mathrm{col}(w_1, w_2, \ldots, w_{\mathtt{w}})$; the behavior $\mathfrak{B}$ is a linear subspace of $\mathfrak{C}^\infty(\mathbb{R},\mathbb{R}^{\mathtt{w}})$ 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 $R \in \mathbb{R}^{\mathtt{g}\times\mathtt{w}}[\xi]$ such that

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

The set of linear time-invariant differential systems with manifest variable w taking its value in $\mathbb{R}^{\mathtt{w}}$ is denoted by $\mathfrak{L}^{\mathtt{w}}$.

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 $R \in \mathbb{R}^{\mathtt{g}\times\mathtt{w}}[\xi]$ be a polynomial matrix. If the behavior $\mathfrak{B}$ is represented by $R(\frac{\mathrm{d}}{\mathrm{d}t})w = 0$ then we call this a *kernel representation* of $\mathfrak{B}$ and we write $\mathfrak{B} = \ker(R(\frac{\mathrm{d}}{\mathrm{d}t}))$. A kernel representation is said to be *minimal* if every other kernel representation of $\mathfrak{B}$ has at least g rows. A given kernel representation, $R(\frac{\mathrm{d}}{\mathrm{d}t})w = 0$, is minimal if and only if the polynomial matrix R has full row rank.

Let $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ and let $R(\frac{\mathrm{d}}{\mathrm{d}t})w = 0$ be a kernel representation of $\mathfrak{B}$. Assume $\mathrm{rank}(R) < \mathtt{w}$ (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 $w = \mathrm{col}(w_1, w_2, \ldots, w_{\mathtt{w}})$ 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 $\mathtt{m}(\mathfrak{B})$). Once $\mathtt{m}(\mathfrak{B})$ free components are chosen, the remaining $\mathtt{w} - \mathtt{m}(\mathfrak{B})$ components are determined up to a finite-dimensional affine subspace of $\mathfrak{C}^\infty(\mathbb{R},\mathbb{R}^{\mathtt{w}-\mathtt{m}(\mathfrak{B})})$. These are called *outputs*, and the number of outputs is denoted by $\mathtt{p}(\mathfrak{B})$, 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 $\mathtt{m}(\mathfrak{B})$ components of u as inputs, and the $\mathtt{p}(\mathfrak{B})$ 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 $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ 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 $R = (Q \quad P)$, 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 the concept of controllability.

Definition 2.1: A system $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ is *controllable* if for all $w_1, w_2 \in \mathfrak{B}$, there exists a $T \geq 0$ and a $w \in \mathfrak{B}$ such that $w(t) = w_1(t)$ for $t < 0$ and $w(t + T) = w_2(t)$ for $t \geq 0$.

It was shown in [6] 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 $\mathtt{w} \times \mathtt{l}$ polynomial matrix M such that

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

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

A system is called autonomous if for any trajectory its future is completely determined by its past:

Definition 2.2: A system $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ is called *autonomous* if for every $w \in \mathfrak{B}$ we have that $w(t) = 0$ for all $t \leq 0$ implies $w = 0$.

Definition 2.3: A system $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ 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.

It was shown in [6] that if $\mathfrak{B} = \ker(R(\frac{\mathrm{d}}{\mathrm{d}t}))$, 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 $\mathbb{C}^+ = \{\lambda \,|\, \mathrm{Re}(\lambda) \geq 0\}$. Note that a stable behavior is necessarily autonomous. A behavior $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ is called *trivial* if $\mathfrak{B} = \{0\}$. The trivial behavior is the only behavior that is both controllable and autonomous. An autonomous behavior $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ is called *anti-stable* if for all nonzero $w \in \mathfrak{B}$ we have either $\lim_{t\to\infty} w(t) \neq 0$ or $\lim_{t\to\infty} w(t)$ does not exist. It can be shown that if $\mathfrak{B}$ is autonomous and $\mathfrak{B} = \ker(R(\frac{\mathrm{d}}{\mathrm{d}t}))$, then $\mathfrak{B}$ is anti-stable if and only if $R(\lambda)$ has full column rank for all $\lambda \in \mathbb{C}^- = \{\lambda \,|\, \mathrm{Re}(\lambda) < 0\}$. 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 2.4: Let $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ be autonomous. Then there exists a unique autonomous stable $\mathfrak{B}_{\mathrm{stab}} \in \mathfrak{L}^{\mathtt{w}}$, and a unique autonomous anti-stable $\mathfrak{B}_{\mathrm{antistab}} \in \mathfrak{L}^{\mathtt{w}}$ such that $\mathfrak{B} = \mathfrak{B}_{\mathrm{stab}} \oplus \mathfrak{B}_{\mathrm{antistab}}$.

The behaviors $\mathfrak{B}_{\mathrm{stab}}$ and $\mathfrak{B}_{\mathrm{antistab}}$ are called the *stable part* and the *anti-stable part* of $\mathfrak{B}$, respectively. It follows from Theorem 3.2.16 in [6] that all trajectories in $\mathfrak{B}_{\mathrm{stab}}$ are stable Bohl functions and all nonzero trajectories in $\mathfrak{B}_{\mathrm{antistab}}$ are anti-stable Bohl functions.

Often, we encounter behaviors $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ that are neither autonomous nor controllable. The *controllable part* of a behavior $\mathfrak{B}$ is defined as the largest controllable subbehavior of $\mathfrak{B}$. This is denoted by $\mathfrak{B}_{\mathrm{cont}}$. A given $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}}$ can always be decomposed as $\mathfrak{B} = \mathfrak{B}_{\mathrm{cont}} \oplus \mathfrak{B}_{\mathrm{aut}}$, where $\mathfrak{B}_{\mathrm{cont}}$ is the (unique) controllable part of $\mathfrak{B}$, and $\mathfrak{B}_{\mathrm{aut}}$ is a (nonunique) autonomous subbehavior of $\mathfrak{B}$. For details we refer to [6].

To conclude this section, we review some facts on elimination of variables. Let $\mathfrak{B} \in \mathcal{L}^{\mathtt{w}_1+\mathtt{w}_2}$ 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 time-invariant 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 $\mathfrak{B} = \ker\begin{pmatrix} R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & R_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix}$ 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 [6], 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. $(\mathfrak{B}_{\mathrm{cont}})_{w_1} = (\mathfrak{B}_{w_1})_{\mathrm{cont}}$ for all $\mathfrak{B} \in \mathfrak{L}^{\mathtt{w}_1+\mathtt{w}_2}$ (see e.g. Lemma 2.10.4 in [7]).

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

$$\mathcal{N}_{w_1}(\mathfrak{B}) = \{w_1 \mid (w_1, 0) \in \mathfrak{B}\},$$

called the *hidden behavior* of w_1 in $\mathfrak{B}$.

III. ACHIEVABILITY

Definition 3.1: Given a linear time-invariant differential system $(\mathbb{R}, \mathbb{R}^{\mathtt{w}_1+\mathtt{w}_2}, \mathcal{P})$, the *plant*, and another linear time-invariant differential system $(\mathbb{R}, \mathbb{R}^{\mathtt{w}_1+\mathtt{w}_2}, \mathcal{O})$, 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 $\mathtt{w}_2$. 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.

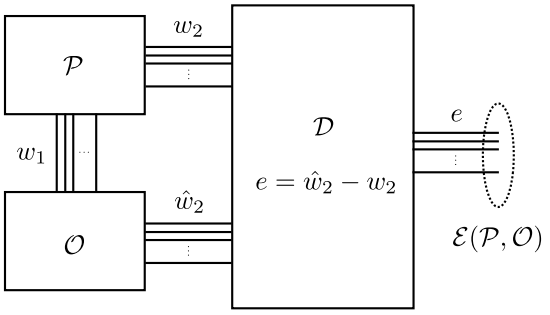

Fig. 1. An observer interconnection gives rise to an error behavior through interconnection with the "differencing system" $\mathcal{D}$.

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

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

where $\mathcal{D} = \{(w_2, \hat{w}_2, e) \,|\, e = \hat{w}_2 - w_2\}$ defines the *error variable* e. The dynamical system $(\mathbb{R}, \mathbb{R}^{\mathtt{w}_2}, \mathcal{E}(\mathcal{P}, \mathcal{O}))$ 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 $\mathcal{P} \in \mathfrak{L}^{\mathtt{w}_1+\mathtt{w}_2}$, a behavior $\mathcal{E} \in \mathfrak{L}^{\mathtt{w}_2}$ 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 $\mathcal{P} = \ker\begin{pmatrix} R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & R_2(\frac{\mathrm{d}}{\mathrm{d}t})\end{pmatrix}$ be a minimal kernel representation and let $\mathcal{O} = \ker\begin{pmatrix} S(\frac{\mathrm{d}}{\mathrm{d}t})R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & S(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t})\end{pmatrix}$, where S is a polynomial matrix. Then $\mathcal{E}(\mathcal{P},\mathcal{O}) = \ker(S(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t}))$.

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

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

Using unimodular row transformations this can be equivalently expressed as

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

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 $\big((\mathcal{P} \wedge_{w_1} \mathcal{O}) \wedge_{(w_2,\hat{w}_2)} \mathcal{D}\big)_e$ is given by the third block row on the right, i.e. $\mathcal{E}(\mathcal{P},\mathcal{O}) = \ker(S(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t}))$. ∎

Proposition 3.5: Let $\mathcal{P} \in \mathfrak{L}^{\mathtt{w}_1+\mathtt{w}_2}$ and let $\mathcal{E} \in \mathfrak{L}^{\mathtt{w}_2}$. Then $\mathcal{E}$ 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: Assume that $\mathcal{E}$ is achieved by $\mathcal{O}$ then $\mathcal{E}(\mathcal{P},\mathcal{O}) = \mathcal{E}$. Let $w_2 \in \mathcal{N}_{w_2}(\mathcal{P})$ be arbitrary. Define $w_1 = 0$ then $(w_1, w_2) \in \mathcal{P}$. Define $\hat{w}_2 = 0$ then $(w_1, \hat{w}_2) = (0,0) \in \mathcal{O}$ and hence $(w_1, w_2, \hat{w}_2, -w_2) \in (\mathcal{P} \wedge_{w_1} \mathcal{O}) \wedge_{(w_2,\hat{w}_2)} \mathcal{D}$ where $e = \hat{w}_2 - w_2 = -w_2$. It follows that $w_2 \in \mathcal{E}(\mathcal{P},\mathcal{O}) = \mathcal{E}$ and hence $\mathcal{N}_{w_2}(\mathcal{P}) \subset \mathcal{E}$.

Conversely, assume that $\mathcal{N}_{w_2}(\mathcal{P}) \subset \mathcal{E}$. Let $\mathcal{P} = \ker\begin{pmatrix} R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & R_2(\frac{\mathrm{d}}{\mathrm{d}t})\end{pmatrix}$ 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 $\mathcal{O} = \ker\begin{pmatrix} S(\frac{\mathrm{d}}{\mathrm{d}t})R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & S(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t})\end{pmatrix}$. 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 explicitely 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 $\mathcal{O}$.

Remark 3.7: The whole theory presented in this paper could easily be extended to include a third subset of *irrelevant* variables w_3 in the plant, cf. [2]. All our results immediately generalize by applying them to the projected plant behavior after elimination of w_3. See Section VI for an example application of this technique.

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], [8].

Definition 4.1: Given a linear time-invariant differential system $(\mathbb{R}, \mathbb{R}^{\mathtt{w}_1+\mathtt{w}_2}, \mathcal{P})$, the variable w_2 is

(1) *observable from* w_1 *(in* $\mathcal{P}$*)* if for all $(w_1, w_2) \in \mathcal{P}$, $w_1 = 0$ implies $w_2 = 0$, i.e. if $\mathcal{N}_{w_2}(\mathcal{P}) = \{0\}$.
(2) *detectable from* w_1 *(in* $\mathcal{P}$*)* if for all $(w_1, w_2) \in \mathcal{P}$, $w_1 = 0$ implies $\lim_{t\to\infty} w_2(t) = 0$, i.e. if $\mathcal{N}_{w_2}(\mathcal{P})$ 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. 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], [8].

Definition 4.2: 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], [8]) are now immediate consequences of these definitions and Proposition 3.5.

Proposition 4.3: Let $\mathcal{P} \in \mathfrak{L}^{\mathtt{w}_1+\mathtt{w}_2}$ 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 $\mathcal{P}$).

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 4.4: Given an observer interconnection, the observer is *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 [9], [10]. 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 4.5: Given a plant $\mathcal{P} \in \mathfrak{L}^{\mathtt{w}_1+\mathtt{w}_2}$ 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 $\mathcal{P} = \ker\begin{pmatrix} R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & R_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix}$ be a minimal kernel representation then $\mathcal{N}_{w_2}(\mathcal{P}) = \ker(R_2(\frac{\mathrm{d}}{\mathrm{d}t}))$. By Proposition 3.5 there exists a polynomial matrix S such that $\mathcal{E} = \ker(S(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t}))$. Since $\mathcal{E}$ is autonomous, SR_2 has full column rank and hence there exists a unimodular matrix U such that

$$USR_2 = \begin{pmatrix} \bar{R} \\ 0 \end{pmatrix}, \tag{1}$$

where $\bar{R}$ is square and nonsingular. It follows that $\mathcal{E} = \ker(\bar{R}(\frac{\mathrm{d}}{\mathrm{d}t}))$. Define

$$\begin{pmatrix} S_1 \\ S_2 \end{pmatrix} = US,$$

where the splitting is as in (1), then $S_1R_2 = \bar{R}$. Define $\mathcal{O} = \ker\begin{pmatrix} S_1(\frac{\mathrm{d}}{\mathrm{d}t})R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & S_1(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix}$. By Lemma 3.4, $\mathcal{E}(\mathcal{P}, \mathcal{O}) = \ker(S_1(\frac{\mathrm{d}}{\mathrm{d}t})R_2(\frac{\mathrm{d}}{\mathrm{d}t})) = \ker(\bar{R}(\frac{\mathrm{d}}{\mathrm{d}t})) = \mathcal{E}$. ∎

The previous proposition implies that we can augment any of the statements in Proposition 4.3 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 4.6: 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 $\mathcal{B}_1 \in \mathfrak{L}^{\mathtt{w}_1}$ be an autonomous behavior and let $\mathcal{B}_2 \in \mathfrak{L}^{\mathtt{w}_1+\mathtt{w}_2}$ 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 representation $\mathcal{B}_1 = \ker(R(\frac{\mathrm{d}}{\mathrm{d}t}))$ where R has full column rank. w_2 being trackable from w_1 (in $\mathcal{B}_2$) implies that $\mathcal{N}_{w_2}(\mathcal{P})$ is autonomous and hence that $\mathcal{B}_2$ admits a kernel representation $\mathcal{B}_2 = \ker\begin{pmatrix} R_1(\frac{\mathrm{d}}{\mathrm{d}t}) & R_2(\frac{\mathrm{d}}{\mathrm{d}t}) \end{pmatrix}$ 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 is nonintrusive. Let

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

be 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 $M_1 = L_1S$.

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 $(\mathcal{P}_{w_1})_{\mathrm{cont}} \subset (\mathcal{O}_{w_1})_{\mathrm{cont}}$ and hence $\mathrm{im}(M_1(\frac{\mathrm{d}}{\mathrm{d}t})) = (\mathcal{P}_{\mathrm{cont}})_{w_1} = (\mathcal{P}_{w_1})_{\mathrm{cont}} \subset (\mathcal{O}_{w_1})_{\mathrm{cont}} = (\mathcal{O}_{\mathrm{cont}})_{w_1} = \mathrm{im}(L_1(\frac{\mathrm{d}}{\mathrm{d}t}))$. The statement now follows as in the proof of Theorem 7.3 in [11]. ∎

For the special case of i/o-observers, the following theorem was previously announced in [12].

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, $\mathcal{P}_{\mathrm{cont}} \subset \mathcal{O}$.

Proof: Assume that $\hat{w}_2$ is trackable from w_1 (in $\mathcal{O}$) and that $\mathcal{P}_{\mathrm{cont}} \subset \mathcal{O}$. Consider the behavior $(\mathcal{P}_{\mathrm{aut}})_{w_1} \wedge_{w_1} \mathcal{O}$, where $(\mathcal{P}_{\mathrm{aut}})_{w_1}$ is the projection of some autonomous part $\mathcal{P}_{\mathrm{aut}}$ of $\mathcal{P}$ onto the variable w_1. By Lemma 5.1, the projection $((\mathcal{P}_{\mathrm{aut}})_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2}$ 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 $(w_1, w_2) =$

$(w_1^{\rm c}, w_2^{\rm c})+(w_1^{\rm a}, w_2^{\rm a})$, where $(w_1^{\rm c}, w_2^{\rm c}) \in \mathcal{P}_{\rm cont}$ and $(w_1^{\rm a}, w_2^{\rm a}) \in \mathcal{P}_{\rm aut}$. Since $\mathcal{P}_{\rm cont} \subset \mathcal{O}$ we have $(w_1, \hat{w}_2) - (w_1^{\rm c}, w_2^{\rm c}) \in \mathcal{O}$. The latter equals $(w_1^{\rm a}, \hat{w}_2 - w_2 + w_2^{\rm a}) = (w_1^{\rm a}, e + w_2^{\rm a})$. Since this is in $\mathcal{O}$, we have $e + w_2^{\rm a} \in ((\mathcal{P}_{\rm aut})_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2}$ and $e \in ((\mathcal{P}_{\rm aut})_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2} + (\mathcal{P}_{\rm aut})_{w_2}$ 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 4.6. Let

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

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

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

Since $\mathcal{P}_{\rm cont} \subset \mathcal{P}$ and $\mathcal{O}_{\rm cont} \subset \mathcal{O}$, it follows that $(\mathcal{P}_{\rm cont} \wedge_{w_1} \mathcal{O}_{\rm cont}) \wedge_{(w_2, \hat{w}_2)} \mathcal{D} \subset (\mathcal{P} \wedge_{w_1} \mathcal{O}) \wedge_{(w_2, \hat{w}_2)} \mathcal{D}$ and hence that $\mathcal{E}(\mathcal{P}_{\rm cont}, \mathcal{O}_{\rm cont}) \subset \mathcal{E}(\mathcal{P}, \mathcal{O})$. But then $\mathcal{E}(\mathcal{P}_{\rm cont}, \mathcal{O}_{\rm cont})$ is autonomous and has output cardinality $\mathtt{w}_2$. Hence

$$\operatorname{rank} \begin{pmatrix} 0 & M_1 & -L_1 \\ -I & M_2 & -L_2 \end{pmatrix} - \operatorname{rank} \begin{pmatrix} M_1 & -L_1 \\ M_2 & -L_2 \end{pmatrix} = \mathtt{w}_2. \quad (2)$$

By Lemma 5.2 there exists a rational matrix S such that $M_1 = L_1 S$. But then

$$\begin{aligned} \operatorname{rank} \begin{pmatrix} 0 & M_1 & -L_1 \\ -I & M_2 & -L_2 \end{pmatrix} &= \operatorname{rank} \begin{pmatrix} 0 & L_1 S & -L_1 \\ -I & M_2 & -L_2 \end{pmatrix} \\ &= \operatorname{rank} \begin{pmatrix} 0 & 0 & -L_1 \\ -I & M_2 - L_2 S & -L_2 \end{pmatrix} \\ &= \operatorname{rank} \begin{pmatrix} 0 & 0 & -L_1 \\ -I & 0 & -L_2 \end{pmatrix} \\ &= \mathtt{w}_2 + \operatorname{rank} \begin{pmatrix} -L_1 \\ -L_2 \end{pmatrix}, \end{aligned}$$

where the last equality follows since our image representation has full column rank. Combining this with (2) yields

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

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 $Q(\frac{\rm d}{{\rm d}t})$ is surjective. This implies that

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

■

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 $\mathcal{B}_1 \in \mathfrak{L}^{\mathtt{w}_1}$ be a stable behavior and let $\mathcal{B}_2 \in \mathfrak{L}^{\mathtt{w}_1+\mathtt{w}_2}$ 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 $\mathcal{B}_1 = \ker(R(\frac{\rm d}{{\rm d}t}))$ 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 $\mathcal{B}_2 = \ker \begin{pmatrix} R_1(\frac{\rm d}{{\rm d}t}) & R_2(\frac{\rm d}{{\rm d}t}) \end{pmatrix}$ where $R_2(\lambda)$ has full column rank for all $\lambda \in \mathbb{C}^+$. But then

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

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 anti-stable 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 $\mathcal{O} = \ker \begin{pmatrix} \hat{R}_1(\frac{\rm d}{{\rm d}t}) & \hat{R}_2(\frac{\rm d}{{\rm d}t}) \end{pmatrix}$ be a kernel representation. Then $\hat{R}_1(\frac{\rm d}{{\rm d}t}) w_1 = -\hat{R}_2(\frac{\rm d}{{\rm d}t}) \hat{w}_2$ 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 $\hat{R}_1(\frac{\rm d}{{\rm d}t}) w_1 = 0$ 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 $\mathcal{P} = \ker \begin{pmatrix} R_1(\frac{\rm d}{{\rm d}t}) & R_2(\frac{\rm d}{{\rm d}t}) \end{pmatrix}$ be a kernel representation, then there exists a polynomial matrix S such that $\hat{R}_1 = S R_1$.

Assume now that w_2 is an anti-stable Bohl function (the alternative case). Using the above kernel representations it follows that $\hat{R}_2(\frac{\rm d}{{\rm d}t}) e = \hat{R}_2(\frac{\rm d}{{\rm d}t}) \hat{w}_2 - \hat{R}_2(\frac{\rm d}{{\rm d}t}) w_2 = -S(\frac{\rm d}{{\rm d}t}) R_1(\frac{\rm d}{{\rm d}t}) w_1 - \hat{R}_2(\frac{\rm d}{{\rm d}t}) w_2 = (S(\frac{\rm d}{{\rm d}t}) R_2(\frac{\rm d}{{\rm d}t}) - \hat{R}_2(\frac{\rm d}{{\rm d}t})) w_2$ 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 $\hat{R}_2(\frac{\rm d}{{\rm d}t})(\hat{w}_2 - w_2) = 0$ 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, a controllable/autonomous decomposition $\mathcal{P} = \mathcal{P}_{\rm cont} \oplus \mathcal{P}_{\rm aut}$ and the associated anti-stable/stable decomposition $\mathcal{P}_{\rm aut} = \mathcal{P}_{\rm antistab} \oplus \mathcal{P}_{\rm stab}$, cf.

Proposition 2.4, 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 controllable part of the plant behavior plus the anti-stable part of the autonomous part, $\mathcal{P}_{\rm cont} \oplus \mathcal{P}_{\rm antistab} \subset \mathcal{O}$.

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

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

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 $(w_1', w_2') \in \mathcal{P}_{\rm cont} \oplus \mathcal{P}_{\rm antistab}$ and $(w_1'', w_2'') \in \mathcal{P}_{\rm stab}$. Since $\mathcal{P}_{\rm cont} \oplus \mathcal{P}_{\rm antistab} \subset \mathcal{O}$ 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 $e + w_2'' \in ((\mathcal{P}_{\rm stab})_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2}$ and $e \in ((\mathcal{P}_{\rm stab})_{w_1} \wedge_{w_1} \mathcal{O})_{\hat{w}_2} + (\mathcal{P}_{\rm stab})_{w_2}$ 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 4.6. Since $\mathcal{E}(\mathcal{P},\mathcal{O})$ is autonomous it follows from Theorem 5.3 that $\mathcal{P}_{\rm cont} \subset \mathcal{O}$. Since $\mathcal{P}_{\rm antistab} \subset \mathcal{P}$, $\mathcal{E}(\mathcal{P},\mathcal{O})$ stable implies that $\mathcal{E}(\mathcal{P}_{\rm antistab},\mathcal{O})$ is stable. Moreover, $\mathcal{O}$ is clearly also nonintrusive with respect to $\mathcal{P}_{\rm antistab}$. It follows from Proposition 5.5 that $\mathcal{P}_{\rm antistab} \subset \mathcal{O}$ and hence that $\mathcal{P}_{\rm cont} \oplus \mathcal{P}_{\rm antistab} \subset \mathcal{O}$. ∎

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 $\mathcal{P} = \ker\begin{pmatrix} R_1(\frac{\rm d}{{\rm d}t}) & R_2(\frac{\rm d}{{\rm d}t})\end{pmatrix}$ be a minimal kernel representation and let $\mathcal{O} = \ker\begin{pmatrix} \hat{R}_1(\frac{\rm d}{{\rm d}t}) & \hat{R}_2(\frac{\rm d}{{\rm d}t})\end{pmatrix}$ be any kernel representation. Then there exists a polynomial matrix S such that $\begin{pmatrix}\hat{R}_1 & \hat{R}_2\end{pmatrix} = \begin{pmatrix} SR_1 & SR_2\end{pmatrix}$. 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, $\mathcal{E}(\mathcal{P},\mathcal{O}) = \ker(S(\frac{\rm d}{{\rm d}t})R_2(\frac{\rm d}{{\rm d}t}))$ 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 4.6. 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], [8]. Because of space constraints we only sketch the general idea, and only for the tracking case. Consider a kernel representation $\mathcal{P} = \ker\begin{pmatrix}R(\frac{\rm d}{{\rm d}t})\end{pmatrix}$ and the Smith form $R = U\begin{pmatrix} D & 0\end{pmatrix}V$ where U and V are unimodular and D is a diagonal polynomial matrix. Then the controllable subbehavior is given by $\mathcal{P}_{\rm cont} = \ker\begin{pmatrix}\begin{pmatrix} I & 0\end{pmatrix}V(\frac{\rm d}{{\rm d}t})^{-1}\end{pmatrix}$. By Theorem 5.3, all nonintrusive tracking observers are of the form

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

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. THE STATE SPACE CASE

In this section we provide a link from our results to classical results from state observer theory.

Consider a plant whose full behavior $\mathcal{P}_{\rm full}$ with variables (u, x, y, z) is given by

$$\begin{aligned} \dot{x} &= Ax + Bu,\\ y &= Cx,\\ z &= Vx. \end{aligned} \tag{3}$$

Here, the various matrices are constant matrices. We denote the projection onto the variables (u, y, z) by $\mathcal{P} = (\mathcal{P}_{\rm full})_{(u,y,z)}$. Consider candidate observers for z from (u, y) whose full behavior $\mathcal{O}_{\rm full}$ with variables $(u, y, v, \hat{z})$ is given by

$$\begin{aligned} \dot{v} &= Kv + Ly + Mu,\\ \hat{z} &= Pv. \end{aligned} \tag{4}$$

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 $\mathcal{O} = (\mathcal{O}_{\rm full})_{(u,y,\hat{z})}$.

Proposition 6.1: $\mathcal{P} = (\mathcal{P})_{(u,y,z)} \subset (\mathcal{O})_{(u,y,\hat{z})} = \mathcal{O}$ if and only if there exists a (constant) matrix U such that

$$\begin{aligned} UA - KU &= LC,\\ M &= UB,\\ V &= PU. \end{aligned} \tag{5}$$

In this case, the error dynamics are given by

$$\begin{aligned} \dot{d} &= Kd,\\ e &= Pd, \end{aligned} \tag{6}$$

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

Proof: In [8] it was shown that given equation (5) the map

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

is a (continuous) behavior homomorphism. This map restricts to the inclusion $\mathcal{P} \subset \mathcal{O}$. Conversely, $\mathcal{P} \subset \mathcal{O}$ implies Equation (5) by Theorem 3.9 in [13]. Equation (6) follows from a simple direct calculation using (3), (4) and (5). ■

In the context of observers for single linear functionals of the state, the Sylvester-type Equation (5) can already be found in Luenbergers original paper [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 (cf. e.g. [14]).

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

Theorem 6.2: Let all uncontrollable modes in (3) be unstable and let (P, K) in (4) be observable. Then (4) 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 (5) 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 (3) are unstable, any autonomous complement of the controllable part of $\mathcal{P} = (\mathcal{P}_{\text{full}})_{(u,y,z)}$ is automatically anti-stable. 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 6.1. ■

Theorem 6.2 provides a generalization of a similar result for controllable plants that is considered classical, although the first full proof of the necessity part the authors know of is relatively recent [15]. See also Remark 3.59 and Remark 3.72 in [14] for a short discussion of the classical literature on this topic.

VII. 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 the estimator variable is detectable from the measured one, and the observer contains both the controllable part of the plant and the antistable part of an autonomous 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 the case where the plant is strictly proper and represented in input-state-output form.

VIII. ACKNOWLEDGMENTS

This work was partially supported by the Australian Research Council through the ARC Discovery Project DP0987411 "State Observers for Control Systems with Symmetry".

The SISTA research program of the K.U. Leuven is supported by the Research Council KUL under projects: GOA AMBioRICS, 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 (ICCoS, 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. Luenberger, "Observing the state of a linear system," *IEEE Transactions on Military Electronics*, vol. 8, no. 2, pp. 74–80, 1964.

[2] M. Valcher and J. Willems, "Observer synthesis in the behavioral approach," *IEEE Transactions on Automatic Control*, vol. 44, no. 12, pp. 2297–2307, 1999.

[3] M. Bisiacco, M. Valcher, and J. 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] M. Belur and H. Trentelman, "Stabilization, pole placement, and regular implementability," *IEEE Transactions on Automatic Control*, vol. 47, no. 5, pp. 735–744, 2002.

[5] A. van der Schaft, "Achievable behavior of general systems," *Systems and Control Letters*, vol. 49, no. 2, pp. 141–149, 2003.

[6] J. Polderman and J. Willems, *Introduction to mathematical systems theory: a behavioral approach*. Springer, 1998.

[7] M. Belur, "Control in a behavioral context," Ph.D. dissertation, Rijksuniversiteit Groningen, 2003.

[8] P. Fuhrmann, "Observer theory," *Linear Algebra and its Applications*, vol. 428, no. 1, pp. 44–136, 2008.

[9] I. Blumthaler and U. Oberst, "T-observers," *Linear Algebra and its Applications*, vol. 430, no. 8-9, pp. 2416–2447, 2009.

[10] I. Blumthaler, "Functional T-observers," *Linear Algebra and its Applications*, vol. 432, no. 6, pp. 1560–1577, 2010.

[11] S. Gottimukkala, S. Fiaz, and H. Trentelman, "Equivalence of rational representations of behaviors," *Systems and Control Letters*, vol. 60, no. 2, pp. 119–127, 2011.

[12] J. Trumpf and J. 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.

[13] J. Trumpf, "Observers for linear time-varying systems," *Linear Algebra and its Applications*, vol. 425, no. 2–3, pp. 303–312, 2007.

[14] ——, "On the geometry and parametrization of almost invariant subspaces and observer theory," Ph.D. dissertation, Universität Würzburg, 2002.

[15] P. 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.