Dynamics Reported, Volume 2
Edited by U. Kirchgraber and H. O. Walther

5
Models for Dynamics

Jan C. Willems
Mathematics Institute, University of Groningen

> Ce que l'on conçoit bien s'énonce clairement,
> Et les mots pour le dire arrivent aisément.
> (Boileau, *l'Art Poétique*, 1674)

CONTENTS

INTRODUCTION

The purpose of this paper is to give a tutorial exposition of what we consider to be the basic mathematical concepts in the theory of dynamical systems.

One of the main goals and key functions of mathematical thinking is to provide an effective and unambiguous language for classifying phenomena, for expressing concepts and ideas, for translating into mathematics vague notions and expressions which we borrow from our daily vocabulary. Examples of major mathematical achievements of this type are *probability theory* in formalizing what we mean by '*chance*', *information theory* in quantifying '*amount of information*', and *control theory* in shaping the notion of '*feedback*'. The present paper can be viewed as an attempt to provide a mathematical framework for discussing '*dynamics*' on a general level, that is, without reference to a specific class of (physical, economic, or engineering) examples.

Notwithstanding the fact that dynamical systems as mathematical objects have implicitly been with us at least since the time of Newton and the beginnings of calculus, it was only in the present century that we have seen the emergence of a theory in which the notion of a dynamical system was explicitly put forward as an independent mathematical concept. The founders of this subject (Poincaré, Birkhoff, and others) appeared naturally very much inspired by physics and mechanics, and viewed dynamical systems primarily in the context of the theory of differential equations. For reasons unknown, external forces had at that time faded away as an integral part of the mathematical description of mechanical systems, even though they were (and still are!) very much present in mechanical engineering applications. Consequently, we saw the emergence of flows on manifolds (and their generalizations to infinite-dimensional state spaces and discrete time systems) as the basic concept on which the mathematical foundations of the subject of dynamics were laid out. This development and line of thought has been maintained until the present days.

In this classical concept, a dynamical system is described in terms of the evolution of its state. It is assumed that the state evolves in an autonomous way. With this, we mean that its path depends only on its initial value and on the laws of motion. In situations other than some very well defined mechanical or electrical systems, the theory leaves us guessing as to how the state variables should be chosen. Further, no external influences are formally incorporated in this framework: the state evolves purely on the basis of internal driving forces. By assuming that the state evolves in this deterministic fashion we postulate in effect that the system is isolated from its environment. But there is no such thing as an isolated system! What this assumption actually means is that we postulate that we know, or that we think to know, how the environment will act on the system, what the boundary conditions are, how external influences are generated—and so, in modelling a specific, concrete, dynamical system in the language of classical dynamics, *we find ourselves in the absurd situation of having to model also the environment*!

A rather independent route to the concept of a dynamical system has been followed in electrical engineering, in particular in areas as circuit theory, control, signal processing, and, later on, computer science. In these fields, particularly in control, there has always been the tendency to view a dynamical system as a '*black box*' which receives stimuli (inputs) from its environment and reacts to these stimuli by producing outputs. The work of Kalman and the other innovations of the 1960s in control theory showed how to incorporate state variables in this input/output framework. This has led to the state space theory of dynamical systems with inputs and outputs.

What really constitutes a dynamical system? How should one conceptualize it? What are the essential common features in the mathematical models for dynamical phenomena? What is a suitable paradigm on which we can base our definitions and, from there, our problem formulations?

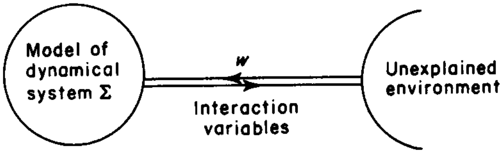


Fig. 1

Our formalization of a dynamical system stems from the mental picture shown in Fig. 1. We view a dynamical system as an object which is imbedded in its environment, is abstracted from it, but which may, will, interact with it. The system has certain attributes whose evolution in time we wish to describe. In order to do so we select the relevant set of time instances, T, and the set W in which the attributes take on their values. The dynamical laws specifying this time evolution tell us that certain trajectories can occur, that others cannot. This yields what we call the *behaviour* of the dynamical system. This point of view takes the model equations, any set of dynamical relations, as basic and proceeds from there. That is what the modeller gives us, that is what a mathematical theory of dynamics should start with. If preconditioning of the model is necessary (for example, in order to display the evolution of the state or the input/output structure), then a theory should make clear how and why this should be done.

Note that it is only by viewing also the environment as a specific dynamical system that we are able to obtain an autonomous classical dynamical system. It is in this sense that classical dynamics forces us to model also the environment (see Fig. 2).

However, most equations which we obtain from first principles will contain other variables in addition to the attributes whose behaviour we are trying to describe. We will call these additional variables *latent variables.* A special class of such latent variables aims at extracting the memory of a system and leads to the notion of the *state* of a dynamical system.

The theme of the first section of this paper is to formalize all this and to illustrate the suitability of our starting point by means of a series of examples taken from a variety of scientific disciplines.

In our philosophy the state of a system is a mathematical object which is of

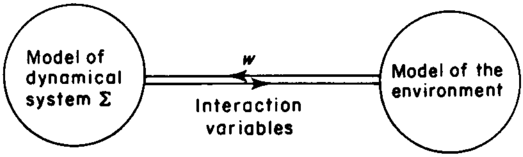


Fig. 2

crucial importance for the analysis and synthesis of dynamical systems. However, seldomly will the state be explicitly part of a mathematical specification of a dynamical system described by equations obtained from first principles. The state and the state equations will have to be deduced from the mathematial model, from a given set of dynamical equations. Writing state equations for dynamical systems is the problem of *realization theory* and is the subject of section 2. The upshot of this development is the fact that the belief that the state of a dynamical systems is uniquely defined *up to relabelling* and *barring hidden variables* is, in general, not true. It is, however, correct for important classes of systems, in particular for linear and for autonomous dynamical systems.

In section 3 we will show how cause/effect—input/output structures emerge in this setting. This class of systems, of great importance in control and signal processing, allows to view the external variables as consisting of *inputs*, stimuli, formalizing the influence of the environment on the system and *outputs*, responses, formalizing the influence of the system on the environment.

In section 4 we will study linear systems in detail. We will show that such systems admit a variety of convenient representations. We will put these representations in the common perspective of what we will call $(ARMA)$ *models*, linear shift invariant relations involving the external attributes and latent variables, and emphasize the role which controllability plays in the representation of such systems.

In his influential book *The Structure of Scientific Revolutions*, Thomas Kuhn argues that a field of scientific inquiry is made up by *paradigms* and *puzzles*. He describes paradigms as models for research, a general problem area sharing a common formulation, a framework in which it becomes possible to ask *'valid'* questions. Puzzles are concrete applications, conjectures, open problems. Most scientists piece together puzzles and it is this activity which Kuhn calls *normal science*. The term puzzle suggests *spielerei*—playing games. This negative connotation is—so it is said—unintentional. By formulating puzzles, a scientist can focus on specific questions, questions lead to answers, answers are the products of scientific research.

This structure of scientific inquiry is very much present in (applied) mathematics in general and in the theory of dynamical systems in particular. However, there has been an unfortunate unexplainable total domination of puzzle solving. Paradigms have been muted, suppressed, not spoken about, let alone scrutinized, rejected, updated. Examining and formulating paradigms has achieved a reputation in mathematical circles as being soft: it leads to too many definitions and not enough theorems. Good mathematics is thought to have a high theorem-to-definition ratio. Solving puzzles, on the other hand, is considered a serious activity, requiring intelligence, mathematical culture, virtuosity. The ultimate of mathematical achievement is to solve a puzzle (a conjecture) formulated by someone else preferably in another century. Thus we have attained a complete reversal in which posing paradigms is considered *spielerei*, we find ourselves in a situation in which proving theorems, not building theories, appears to be the aim of mathematical research.

It is, therefore, with a certain amount of hesitation that the present paper has been written. Its purpose is to present the formalization of the picture of Fig. 1 as the paradigm of a dynamical system in the hope of showing its usefulness in mathematics, engineering, and physics alike.

1 MODELS FROM FIRST PRINCIPLES

In the first section of this paper, we will introduce a mathematical definition of the concept of a *dynamical system*. We will take the point of view that a dynamical system consists of (a family of) laws which constrain the signals which the system can conceivably produce. The collection of all the signals compatible with these laws define what we call the *behaviour* of the system. However, laws and models which we write down from first principles will invariably contain, in addition to the variables which are being modelled, also other variables: we will call these *latent variables*. Some latent variables may have important properties related to and capturing the memory structure of a system. This leads in particular to the concept of the *state* of a dynamical system. We will illustrate the abstract concepts introduced in this section by means of a series of concrete examples. We will also show how systems described by difference or differential equations fit in our abstract setting and finally how our concepts are related to the notions of formal languages and automata.

1.1 The notion of a dynamical system—examples

1.1.1 *The basic concept*

Let us start at the very beginning: with our definition of a dynamical system. We will use this definition as a *leitmotiv* throughout this paper. The definition is hopelessly general but nevertheless it captures rather well the crucial features of the notion of a dynamical system.

DEFINITION 1.1 A *dynamical system* Σ is defined as a triple

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

with $T \subseteq \mathbb{R}$ the *time axis*; W an abstract set, called the *signal alphabet*; and $\mathfrak{B} \subseteq W^T$ *the behaviour*.

The set T specifies the set of time instances relevant to our problem. Usually $T = \mathbb{R}$ or $\mathbb{R}_+$ (in continuous time systems), $\mathbb{Z}$ or $\mathbb{Z}_+$ (in discrete time systems), or, more generally, an interval in $\mathbb{R}$ or $\mathbb{Z}$. We view a dynamical system as an entity which is abstracted from its environment but which interacts with it (see Fig. 1). The set W specifies the way in which the attributes of the dynamical

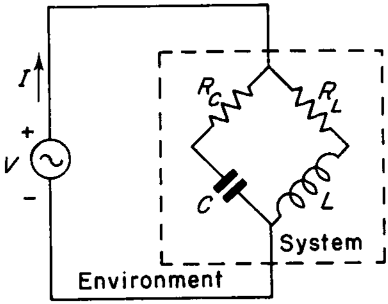


Fig. 3

system are formalized as elements of a set. These attributes are the variables whose evolution in time we are describing. These will be a combination of observed variables and variables through which the system interacts with its environment. (If we think of the observer and the modeller as being part of the environment then we can consider the description of this interaction with the environment as the essential feature of the attributes.)

The behaviour $\mathfrak{B}$ is simply a family of time trajectories taking their values in the signal alphabet. Thus elements of $\mathfrak{B}$ constitute precisely the trajectories compatible with the laws which govern the system: $\mathfrak{B}$ consists of all time signals which—according to the model—our system can conceivably generate. In most applications, the behaviour $\mathfrak{B}$ will be specified by equations, often differential or difference equations, sometimes integral equations. In other words, there is a map $b: W^{\mathrm{T}} \rightarrow E$ with $E = \{0, 1\}$, or more generally a vector space such that $\mathfrak{B} = b^{-1}(0)$. We will call such equations *behavioural equations.*

1.1.2 *An electrical circuit*

Let us look at a number of typical examples of how dynamical models are constructed from first principles. Our first example considers the terminal behaviour of the electrical circuit shown in Fig. 3.

The circuit interacts with its environment through the external port. The attributes which describe this interaction are the current I into the circuit and the voltage V across its external terminals. Hence $W = \mathbb{R}^2$. As time axis in this example we take $T = \mathbb{R}$. In order to specify the terminal behaviour, we will introduce the currents through and the voltages across the internal branches of the circuit, as shown in Fig. 4.

The following behavioural equations must be satisfied:

Constitutive equations:

$$V_{R_C} = R_C I_{R_C}; \quad V_{R_L} = R_L I_{R_L}; \quad C\dot{V}_C = I_C; \quad L\dot{I}_L = V_L \qquad (EC_1)$$

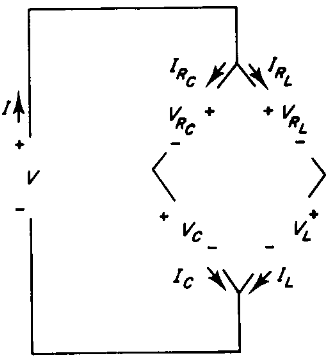


Fig. 4

Kirchhoff's current laws:

$$I = I_{R_C} + I_{R_L};\quad I_{R_C} = I_C;\quad I_{R_L} = I_L \tag{EC_2}$$

Kirchhoff's voltage laws:

$$V = V_C + V_{R_C} = V_L + V_{R_L} \tag{EC_3}$$

This yields the *port behaviour*, formally defined as:

$$\mathfrak{B} = \{(I, V): \mathbb{R} \to \mathbb{R}^2 \mid \exists (I_{R_C}, V_{R_C}, I_{R_L}, V_{R_L}, I_C, V_C, I_L, V_L): \mathbb{R} \to \mathbb{R}^8$$

satisfying equations $(EC)\}$

After elimination of the variables I_{R_C}, V_{R_C}, I_{R_L}, V_{R_L}, I_C, I_L, V_L, we obtain as behavioural equation the differential equation

$$R_C L C \ddot{I} + (L + R_C R_L C)\dot{I} + R_L I = LC\ddot{V} + (R_C + R_L)C\dot{V} + V \tag{EC'}$$

This yields the following explicit specification of the behaviour

$$\mathfrak{B} = \{(I, V): \mathbb{R} \to \mathbb{R}^2 \mid (EC') \text{ is satisfied}\}$$

In both the above specifications of $\mathfrak{B}$ we have been vague about the precise smoothness conditions required of the signals. This issue is not especially important to us at the moment.

1.1.3 *A Leontieff economy*

As a second example let us consider a Leontieff model for an economy in which several economic goods are transformed by means of a number of production processes. We are interested in describing the evolution in time of the total utility of the goods in the economy. Assume that there are N production processes in which n economic goods are transformed into goods of the same kind, and that in order to produce one unit of good j by means of the kth production process, we need at least a_{ij}^k units of good i. The real numbers a_{ij}^k, $k \in \boldsymbol{N}$, $i, j \in \boldsymbol{n}, := \{1, 2, \ldots, n\}$ are called the *technology coefficients*. We assume that in each time unit one production cycle will take place.

Denote by

$q_i(t)$ the quantity of product i available at time t;
$u_i^k(t)$ the quantity of product i assigned to the production process k at time t;
$y_i^k(t)$ the quantity of product i acquired from the production process k at time t.

There holds:

$$\sum_{k=1}^{n} u_i^k(t) \leq q_i(t) \qquad \forall i \in \mathbf{n} \qquad (LE_1)$$

$$\sum_{j=1}^{n} a_{ij}^k y_j^k(t+1) \leq u_i^k(t) \qquad \forall k \in N, \quad i \in \mathbf{n} \qquad (LE_2)$$

$$q_i(t+1) \leq \sum_{k=1}^{n} y^k(t+1) \quad \forall i \in \mathbf{n} \qquad (LE_3)$$

The underlying structure of this economy is shown in Fig. 5. The difference between the right- and the left-hand sides of the above inequalities will be due to such things as inefficient production, imbalance of the available products, consumption, and other forms of waste.

Now assume that the total utility of the goods in the economy is a function of the available amount of goods $q_1, q_2, \ldots, q_n$, i.e., $J:\mathbb{Z} \to \mathbb{R}_+$ is given by $J(t) = \eta(q_1(t), \ldots, q_n(t))$, with $\eta:\mathbb{R}_+^n \to \mathbb{R}_+$ a given function, the *utility*. For example, if we identify utility with resale value (in dollars, of course), then $\eta(q_1, q_2, \ldots, q_n)$ will be equal to $\sum_{k=1}^{n} p_i q_i$ with p_i the per unit selling price of good i. These relations define a dynamical system with $T = \mathbb{Z}$, $W = \mathbb{R}_+$, and

$$\mathfrak{B} = \{J:\mathbb{Z} \to \mathbb{R}_+ \mid \exists q_i:\mathbb{Z} \to \mathbb{R}_+, u_i^k:\mathbb{Z} \to \mathbb{R}_+, y_i^k:\mathbb{Z} \to \mathbb{R}_+, \quad i \in \mathbf{n}, \quad k \in N,$$
such that the inequalities (LE) are satisfied for all $t \in \mathbb{Z}$, and $J = \eta_0(q_1, q_2, \ldots, q_n)\}$.

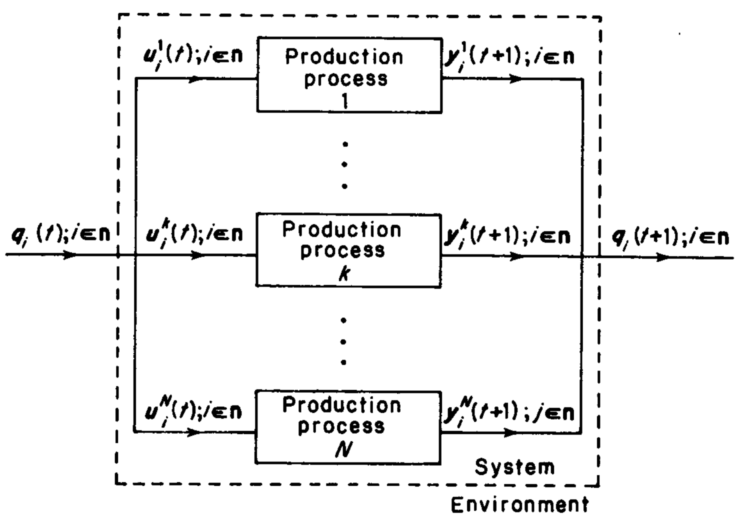


Fig. 5

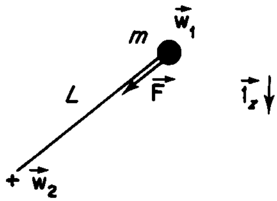


Fig. 6

such that the inequalities (Ec) are satisfied for all $t \in \mathbb{Z}$, and $J = \eta_0(q_1, q_2, \ldots, q_n)\}$. Note that, in contrast with Example 1.1.3 where it was easy to obtain behaviour equations (EC') explicitly in terms of the external attributes V and I, it will be all but impossible in the present example to eliminate the q's, u's and y's and obtain an explicit behavioural equation describing $\mathfrak{B}$ entirely in terms of J, the attribute of interest in this example.

1.1.4 *A pendulum*

Consider the pendulum shown in Fig. 6. Assume that we want to model the relationship between the position $\vec{w}_1$ of the mass and the position $\vec{w}_2$ of the tip of the pendulum (say with the ultimate goal of designing a controller which stabilizes $\vec{w}_1$ at a fixed value by using $\vec{w}_2$ as control, as we do when we balance an inverted broom on our hand). In order to obtain such a model, introduce as auxiliary variables the force $\vec{F}$ in the bar and the real-valued proportionality factor a of $\vec{F}$ and $\vec{w}_1 - \vec{w}_2$.

We obtain the behavioural equations

$$m\frac{d^2\vec{w}_1}{dt^2} = mg\vec{1}_z + \vec{F} \tag{P1}$$

$$\|\vec{w}_1 - \vec{w}_2\| = L \tag{P2}$$

$$\vec{F} = a(\vec{w}_1 - \vec{w}_2) \tag{P3}$$

Here m denotes the mass of the pendulum, L its length, g the gravitational constant, and $\vec{1}_z$ the unit vector in the z-direction. The above equations completely specify the behaviour, defined as

$$\mathfrak{B} = \{(\vec{w}_1, \vec{w}_2): \mathbb{R} \to \mathbb{R}^3 \times \mathbb{R}^3 \mid \exists \vec{F}: \mathbb{R} \to \mathbb{R}^3 \text{ and } a: \mathbb{R} \to \mathbb{R}$$

such that equations (P) are satisfied.$\}$.

1.1.5 *Latent variables*

We have already mentioned that the behaviour of a dynamical system is often specified by *behavioural equations*, that is, there is a map $b: W^{\mathrm{T}} \to \{0, 1\}$ and $\mathfrak{B} := b^{-1}(0)$. On the other hand, one could specify how elements of $\mathfrak{B}$ are *produced*: there is a map $p: P \to W^{\mathrm{T}}$ and $\mathfrak{B} := \operatorname{im} p$. Most common, however, is the situation in which a behaviour is specified as a *projection*. We will now explain this.

As can be observed from the previous examples, models which we write down from first principles will invariably involve, in addition to the basic variables which we are trying to describe, also auxiliary variables (for example—the internal voltages and currents in the electrical circuit of Example 1.12; the q's u's and y's in the economy of Example 1.1.3; the force $\vec{F}$ in the bar of the pendulum and the proportionality factor a in Example 1.1.4). We will call such variables *latent variables.* These latent variables could be introduced, if for no other reason, because they make it more convenient to write down the equations of motion, or because they are essential in order to express the constitutive laws or the conservation laws defining the system's behaviour. This was the case in Examples 1.1.2 and 1.1.4. Latent variables will unavoidably occur whenever we model a system by '*tearing*' where we view a system as an interconnection of subsystems—a common and very useful way of constructing models. This was the case in Example 1.1.4. After interconnection, the external variables of the subsystems will become latent variables for the interconnected system.

Latent variables also play an important role in theoretical considerations. We will see later that latent variables, as state variables or free driving variables, are needed and make it possible to reduce equations of motion to expressions which are purely local in time.

In thinking about the difference between signal variables and latent variables it is helpful *in first instance* to think of the signal variables being directly observable: they are explicit, while the latent variables are not: they are implicit. Examples: in pedagogy, scores of tests can be viewed as the signal, and native intelligence can be viewed as a latent variable aimed at explaining these scores. In thermodynamics, pressure, temperature and volume can be viewed as the signal while the internal energy and entropy can be viewed as latent variables whose value, if needed, should be deduced from the signal. In economics, sales can be viewed as signals, while consumer demand could be considered as a latent variable. We emphasize, however, that which variables are observed and measured, and which are not, is really something which is part of the instrumentation and the technological set-up for a system. Particularly in control applications one should not be *cavalier* about declaring certain variables measurable and observed.

The formalization of systems with latent variables leads to the following definition.

DEFINITION 1.2 A *dynamical system with latent variables* is a quadruple

$$\Sigma_a = (T, W, A, \mathfrak{B}_a)$$

with T, W as in Definition 1.1; *A the set of **latent variables*** and $\mathfrak{B}_a \subseteq (W \times A)^T$ the (extended) behaviour.

Define $P_w: W \cdot\times A \to W$ by $P_w(w, a) := w$. We will call Σ_a a *model with latent variables for the induced dynamical system* $\Sigma = (T, W, P_w\mathfrak{B}_a)$. Often we will refer to, and think of, $\mathfrak{B}_a$ as the *internal* behaviour and of $P_w\mathfrak{B}_a$ as the *external* behaviour of the system.

Note our slight abuse of notation. Whereas in principle $P_w: W \times A \to W$, we also consider it as a map $P_w:(W \times A)^T \to W^T$, yielding $(P_w w)(t) = P_w(w(t))$.

Let us briefly summarize our modelling language in a set theoretic setting. Assume that we have a *phenomenon*, described by *attributes*. We formalize the situation by considering the attributes to belong to a *universum* $\mathfrak{U}$. A mathematical model is a subset $\mathfrak{M}$ of $\mathfrak{U}$: it says that, according to the model, only attributes in $\mathfrak{M}$ can occur and the others cannot. The set $\mathfrak{M}$ is called the *behaviour* of the model. In a model with *latent variables*, we introduce auxiliary variables whose attributes belong to a set $\mathfrak{L}$. A latent variable model is then a subset $\mathfrak{M}^f$ of $\mathfrak{U} \times \mathfrak{L}$. We call $\mathfrak{M}^f$ the *full behaviour*. It induces the (*intrisic*) *behaviour* $\mathfrak{M} = \{u \in \mathfrak{U} \mid \exists l \in \mathfrak{L} \in (u, l) \in \mathfrak{M}^f\}$.

Often behaviours are specified by *behavioural equations*: there is a map $b:\mathfrak{U} \to \{0, 1\}$ and $\mathfrak{M} = b^{-1}(0)$. However, behaviours can also be specified as images, in which case it is logical to consider the domain as the space of latent variables: there is a map $p:\mathfrak{L} \to \mathfrak{U}$, $\mathfrak{M}^f$ is the graph of p, and $\mathfrak{M} = \operatorname{im} p$. Most general, however, the situation in which $\mathfrak{M}$ is the projection of $\mathfrak{M}^f$ and $\mathfrak{M}^f$ itself is described by the *full behavioural equations* $b^f(u, l) = 0$. Inequalities, as in Example 1.3, also occur, however.

1.2 Basic structure

One of the advantages of making definitions at the level of generality of Definition 1.1 is that standard mathematical structures become immediately applicable to dynamical systems.

1.2.1 *Linearity*

We will call the dynamical system $\Sigma = (T, W, \mathfrak{B})$ *linear* if W is a vector space and $\mathfrak{B}$ is a linear subspace of W^{T} (viewed as a vector space in the natural way by means of pointwise addition and scalar multiplication). Example 1.2 is an example of a linear system.

1.2.2 *Time invariance*

We will call the dynamical system $\Sigma = (T, W, \mathfrak{B})$ *time-invariant* if T is an additive semigroup in $\mathbb{R}$ (i.e., $\{t_1, t_2 \in T\} \Rightarrow \{t_1 + t_2 \in T\}$) and $\sigma^t\mathfrak{B} \subseteq \mathfrak{B}$ for all $t \in T$; σ^t-denotes the *backwards or left t-shift*: $(\sigma^t f)(t') := f(t' + t)$. Examples 1.2, 1.3 and 1.4 are time invariant. Example 1.3 can be made time varying in a natural way by assuming that the technology coefficients depend explicitly on time in order to reflect such things as ageing of the machine park and technological progress. Alternatively, we can let the prices p_i be time dependent.

1.2.3 *Symmetry*

Let Σ be a family of dynamical systems: each element of Σ is a dynamical system as in Definition 1.1. Let $\mathscr{G}$ be a group and $\mathfrak{S} = (S_g, g \in \mathscr{G})$ be a transformation group on Σ, that is, each $S_g:\Sigma \to \Sigma$ is a bijection with $S_{g_1 \circ g_2} = S_{g_1} \circ S_{g_2}$. We will call $(\Sigma, \mathfrak{S})$ a *symmetry structure*. An element $\Sigma \in \Sigma$ is said to be ($\mathfrak{S}$-)*symmetric* if $S_g\Sigma = \Sigma$ for all $g \in \mathcal{G}$. Informally, we will say that Σ *has* $\mathfrak{S}$ *as a symmetry*.

Examples of such symmetries are:

(i) Take $T = \mathcal{G}$ to be an additive subgroup of $\mathbb{R}$ and $S_g(T, W, \mathfrak{B}) = (T + g, W, \sigma^g \mathfrak{B})$. The symmetric systems in this sense are, in fact, the time invariant ones.
(ii) Let $(S_g^w, g \in \mathcal{G})$ be a transformation group on W and $S_g(T, W, \mathfrak{B}) = (T, W, S_g\mathfrak{B})$, where, as before and in the sequel, we use the notation $S_g\mathfrak{B} = \{S_g(w(\cdot)): T \to W \mid w \in \mathfrak{B}\}$. The resulting symmetry suggests a behaviour which is invariant under certain sign changes or permutations of the components of the external variables as, for example, a permutation of particles in n-particle systems with identical particles.
(iii) Take $\mathcal{G} = \{0, 1\}$, and define $S_1(T, W, \mathfrak{B}) = (-T, W, R\mathfrak{B})$ where R is the *time reversal*: $(Rf)(t) := f(-t)$. The resulting symmetric systems are called *time reversible*. Examples of time reversible systems are systems described by differential equations containing only even order derivatives.
(iv) Let J be an involution on W (i.e. $J = J^{-1}$). Take $\mathcal{G} = \{0, 1\}$, and define $S_1(T, W, \mathfrak{B}) = (-T, W, JR\mathfrak{B})$. The resulting symmetry is what is sometimes called *dynamic time reversibility*. The involution J serves to express that in order to obtain time reversibility in mechanical systems, it may be necessary to change the sign of the velocities.

1.3 More notation

1.3.1 *Concatenation and non-anticipating maps*

When studying dynamical systems an important role is set aside for the interaction of the past and future of (families of) time functions, for concatenating pasts with futures, and for the way the past and the future interact with maps.

Let $T \subseteq \mathbb{R}$ and W be a set. For a given map $w: T \to W$ we define the following derived maps:

$$
\begin{aligned}
w^- &:= w|_{T \cap (-\infty, 0)} && \text{(the \textit{strict past} of } w) \\
w^{-0} &:= w|_{T \cap (-\infty, 0]} && \text{(the \textit{past and present} of } w) \\
w^+ &:= w|_{T \cap (0, \infty)} && \text{(the \textit{present and future} of } w) \\
w^{0+} &:= w|_{T \cap [0, \infty)} && \text{(the \textit{strict future} of } w)
\end{aligned}
$$

For $\mathfrak{B} \subseteq W^{\mathsf{T}}$, this yields the self-evident notation $\mathfrak{B}^-$, $\mathfrak{B}^{-0}$, $\mathfrak{B}^{0+}$, and $\mathfrak{B}^+$.

Let $w_1, w_2: T \to W$ and $t \in T$. We define the *concatenation at t* of w_1 and w_2, $w_1 \underset{t^-}{\wedge} w_2$ and $w_1 \underset{t^+}{\wedge} w_2$, both maps from T to W, as follows

$$
\left(w_1 \underset{t^-}{\wedge} w_2\right)(t') := \begin{cases} w_1(t') & \text{for } t' < t \\ w_2(t') & \text{for } t' \geq t \end{cases}
$$

$$
\left(w_1 \underset{t^+}{\wedge} w_2\right)(t') := \begin{cases} w_1(t') & \text{for } t' \leq t \\ w_2(t') & \text{for } t' > t \end{cases}
$$

For $\mathfrak{B}_1, \mathfrak{B}_2 \subseteq W^{\mathsf{T}}$ this yields the self-evident notation $\mathfrak{B}_1 \underset{t^-}{\wedge} \mathfrak{B}_2$ and $\mathfrak{B}_1 \underset{t^+}{\wedge} \mathfrak{B}_2$.

We will also concatenate maps which are themselves already restrictions. Thus $w_1^- \underset{0^-}{\Lambda} w_2^{0+} := w_1 \underset{0^-}{\Lambda} w_2$ and $w_1^{-0} \underset{0^+}{\Lambda} w_2^+ := w_1 \underset{0^+}{\Lambda} w_2$, etc. Note that in discrete time ($T = \mathbb{Z}$) there holds $\underset{(t+1)^-}{\Lambda} = \underset{t^+}{\Lambda}$ and as such there would have been no need to introduce both $\underset{t^-}{\Lambda}$ and $\underset{t^+}{\Lambda}$. However, in continuous time we need both.

Let $T \subseteq \mathbb{R}$, W_1 and W_2 be sets, and $\mathfrak{B}_1 \subseteq W_1^T$, $\mathfrak{B}_2 \subseteq W_2^T$. Consider the map $F: \mathfrak{B}_1 \to \mathfrak{B}_2$. We will call F *non-anticipating* if $\{w_1', w_1'' \in \mathfrak{B}_1, t \in T$, and $w_1'(t') = w_1''(t')$ for $t' \leq t\} \Rightarrow \{(Fw_1')(t') = (Fw_1'')(t')$ for $t' \leq t\}$. We will call F *strictly non-anticipating* if instead $\{w_1', w_1'' \in \mathfrak{B}_1, t \in T$, and $w_1'(t') = w_1''(t')$ for $t' < t\} \Rightarrow \{(Fw_1')(t') = (Fw_1'')(t')$ for $t' \leq t\}$. In order to appreciate why in the notion of non-anticipation we use the inequality $t' \leq t$ instead of $t' < t$, consider the following example. Let $T = \mathbb{R}$, $W_1 = W_2 = \mathbb{R}$, $\mathfrak{B}_1 := \{w_1 : \mathbb{R} \to \mathbb{R} \mid w_1$ is bounded and piecewise continuous with $w_1(t) = \lim_{t' \uparrow t} w_1(t')\}$. Now consider the map $F: \mathfrak{B}_1 \to (\mathbb{R})^{\mathbb{R}}$ defined by $(Fw_1)(t) := \lim_{t' \downarrow t} w_1(t)$. Such a map should *not* be non-anticipating but if we would have used $t' < t$ in the definition, it would have been non-anticipating.

An important restriction

Mainly for reasons of exposition and ease of notation *we will* henceworth *restrict attention* in this paper to *time invariant systems with time axis* $T = \mathbb{R}$ *or* $\mathbb{Z}$. Occasionally we will emphasize this assumption in the statement of theorems and definitions.

1.4 The extent of the memory

The memory in a dynamical system, the way the past influences the future, the fact that there is an *after effect*, is what makes dynamical systems interesting, what gives them their idiosyncrasies, what distinguishes them from arbitrary relations and maps. We will now give a series of definitions aimed at classifying the memory structure.

There are four different related angles from which we will look at the memory structure of a dynamical system:

(1) *Completeness*—connected to the possibility of writing the behavioural equations as difference equations;
(2) *The memory span*—expressing the length of the time interval through which the past and future are linked;
(3) *The state*—latent variables which parametrize the content of the system's memory;
(4) *Controllability and autonomy*—notions formalizing in how far the past has a lasting influence on the future.

1.4.1 *Completeness*

The first concept, completeness, requires that the behavioural equations should not extend all the way back to $-\infty$ or all the way forward to $+\infty$. It is a concept which will play an important role in the remainder of our paper. The dynamical system $\Sigma = (T, W, \mathfrak{B})$ is said to be *complete* if

$$\{w \in \mathfrak{B}\} \Leftrightarrow \{w|_{[t_0,t_1]} \in \mathfrak{B}|_{[t_0,t_1]} \text{ for all } -\infty < t_0 \leq t_1 < \infty\}$$

It will be called *L-complete* if

$$\{w \in \mathfrak{B}\} \Leftrightarrow \{w|_{[t,t+L]} \in \mathfrak{B}|_{[t,t+L]} \text{ for all } t \in T\}$$

If a system is L-complete for all $L > 0$, then we will call it *locally specified*. If a system is 0-complete, we will call it *instantly specified*.

The intuitive significance of these notions should be quite obvious. It is clear that a discrete time a dynamical system will be governed by a set of difference equations:

$$f(w(t+L), \quad w(t+L-1), \ldots, w(t)) = 0 \qquad t \in \mathbb{Z}$$

iff it is L-complete. Indeed, take for $f: W^L \to \mathbb{R}$ any map such that $f^{-1}(0) = \mathfrak{B}|_{[0,L]} \in W^L$. (Formally, the above difference equation defines the behaviour $\mathfrak{B} = \{w: \mathbb{Z} \to W|$ the difference equation is satisfied for all $t \in \mathbb{Z}\}$.) We will call the integer $L \in \mathbb{Z}_+$ appearing in this behavioural equation the *lag*. Similarly (disregarding smoothness issues for the time being), a continuous time system governed by a set of differential equations:

$$f\left(\frac{d^n w}{dt^n}(t), \frac{d^{n-1} w}{dt^{n-1}}(t), \ldots, w(t)\right) = 0 \qquad t \in \mathbb{R}$$

will be locally specified. Note that a system is instantly specified iff it is governed by non-dynamic laws, that is, if it is described by behavioural equations of the form: $f(w(t)) = 0$, $t \in T$.

1.4.2 *The memory span*

The dynamical system $\Sigma = (T, W, \mathfrak{B})$ is said to have Δ-finite memory (or equivalently, we say that its *memory span* is Δ) if

$$\{w_1, w_2 \in \mathfrak{B}, \text{ and } w_1|_{[0,\Delta)} = w_2|_{[0,\Delta)}\} \Rightarrow \left\{w_1 \underset{0^-}{\wedge} w_2 \in \mathfrak{B}\right\};$$

it is said to have *finite memory* if it has Δ-finite memory for some $\Delta > 0$; *local memory* if it has Δ-finite memory for all $\Delta > 0$. We will often refer to the memory span when we implicitly mean $\Delta_{\min}$, the *minimal* $\Delta \in \mathbb{Z}_+$ having the above property. When the memory consists of the present value only, we will borrow a notion from the theory of stochastic processes: Σ is said to be *Markovian* if $\{w_1, w_2 \in \mathfrak{B}, w_1(0) = w_2(0)\} \Rightarrow \{w_1 \underset{0^-}{\wedge} w_2 \in \mathfrak{B}\}$. The system is said to be *memoryless* if $\mathfrak{B}$ is closed under concatenation, i.e., if $\{w_1, w_2 \in \mathfrak{B}\} \Rightarrow \{w_1 \underset{0^-}{\wedge} w_2 \in \mathfrak{B}\}$.

Completeness and the memory span are closely related. This is expressed in our first proposition.

PROPOSITION 1.1 Let $\Sigma = (\mathbb{Z}, W, \mathfrak{B})$ be a discrete time dynamical system. Then

$$\{\Sigma \text{ is } t\text{-complete}\} \Leftrightarrow \{\Sigma \text{ is complete and has } t\text{-finite memory}\}$$

Proof ($\Rightarrow$): is obvious.

($\Leftarrow$): Assume that Σ has t-finite memory and that $w: T \to W$ has the property $w|_{[t',t'+t]} \in \mathfrak{B}|_{[t',t'+t]}$ for all $t' \in \mathbb{Z}$. In particular, there exist w_1, w_2, w_3, w_4 such that $w_1 \underset{(t'-1)^-}{\wedge} w \underset{(t'+t)^-}{\wedge} w_2 \in \mathfrak{B}$ and $w_3 \underset{(t')^-}{\wedge} w \underset{(t'+t+1)^-}{\wedge} w_4 \in \mathfrak{B}$. By the t-finite memory assumption, we obtain $w_1 \underset{(t'-1)^-}{\wedge} w \underset{(t'+t+1)}{\wedge} w_4 \in \mathfrak{B}$. Hence $w|_{[t',t'+t+1]} \in \mathfrak{B}|_{[t',t'+t+1]}$ for all $t' \in \mathbb{Z}$. We conclude that $\{w|_{[t',t'+t]} \in \mathfrak{B}|_{[t',t'+t]}\} \Rightarrow \{w|_{[t',t'+t+1]} \in \mathfrak{B}|_{[t',t+t]}\}$. Continuing this process yields $w|_{[t_0,t_1} \in \mathfrak{B}|_{[t_0,t_1]} \forall t_0, t_1 \in \mathbb{Z}$. By completeness, this yields the implication ($\Rightarrow$). ■

We also have the following implications:

instantly specified $\Rightarrow$ locally specified $\Rightarrow$ t-complete $\Rightarrow$ complete;

$\Downarrow$ $\Downarrow$ $\Downarrow$

memoryless $\Rightarrow$ Markovian $\Rightarrow$ local memory $\Rightarrow$ t-finite memory $\Rightarrow$ finite memory.

If $T = \mathbb{Z}$, and if Σ is complete then, by the above proposition, the vertical arrows can also be reversed.

We conclude from all this that a discrete time system can be described by a difference equation with lag L iff it is complete and its memory span is L.

1.4.3 *Splitting variables*

The interaction of the latent variables with the memory structure of a system is of much interest. We will explore this in the next sections. The first concept formalizes situations where the present value of the latent variables all by itself determines the future behaviour of the external signal variable. Let $\Sigma_a = (T, W, A, \mathfrak{B}_a)$ be a dynamical system with latent variables. We will say that the latent variable *splits* the external behaviour if

$$\{(w_1, a_1), (w_2, a_2) \in \mathfrak{B}_a, \quad \text{and} \quad a_1(0) = a_2(0)\} \Rightarrow \left\{ w_1 \underset{0^-}{\wedge} w_2 \in P_w \mathfrak{B}_a \right\}.$$

1.4.4 *State space systems*

If we combine the splitting and the Markov property we arrive at the following very important class of systems.

DEFINITIONS 1.3 Let $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ be a dynamical system with latent variables. We will call this a *dynamical system in state space form*, with *state space*

X, if the behaviour $\mathfrak{B}_s \subseteq (W \times X)^T$ satisfies what we will call the *axiom of state*. This requires the implication $\{(w_1, x_1), (w_2, x_2) \in \mathfrak{B}_s$, and $x_1(0) = x_2(0)\} \Rightarrow \{(w_1, x_1) \underset{0^-}{\wedge} (w_2, x_2) \in \mathfrak{B}_s\}$.

We will call $P_w\mathfrak{B}_s$ the *external behaviour* of Σ_s, and $(T, W, P_w\mathfrak{B}_s)$ the system *induced* by Σ_s. Conversely, we will call $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ a *state space representation* (or a *state space realization*) of $\Sigma = (T, W, P_w\mathfrak{B}_s$. Finally, we will call $\mathfrak{B}_x = P_x\mathfrak{B}_s$ the state behaviour, where $P_x:(W \times X) \to X$ is the projection $P_x(w, x) = x$. On a few occasions we will have the need to consider the states all by themselves, yielding the *state system* $\Sigma_x = (T, X, \mathfrak{B}_x)$.

It is easy to see that in a state space system x splits w, that $(T, W, X, \mathfrak{B}_s)$ is Markovian, and that the *state system* $\Sigma_x = (T, W, \mathfrak{B}_x)$ is also Markovian. However in a state space system x all by itself splits w and x *jointly*. The splitting and the state property will be pursued in detail in section 2. The state space structure will in fact be one of the main issues analysed in the sequel of this paper. Most models which one deals with in physics, economics, dynamic simulation, dynamic control and estimation, etc., are in state space form. We stress, however, that we *do not* view the state as something which is given from first principles but as a variable which should be constructed on the basis of a model given, say, directly in terms of its external behaviour, or in terms of a model incorporating latent variables. It is in this form that mathematical models are obtained from physical or economic principles, and it is from this starting point that a mathematical theory of dynamics should depart. We will discuss the construction of the state space in section 2.

We would like to emphasize that we have assumed that in the splitting or state property the past is the *strict* past and the future contains the present. As such our axiom of state is not quite invariant under time reversal.

1.4.5 *Autonomous, controllable and trim systems*

The notions of autonomous system and controllability aim at classifying in how far the past has lasting implications on the future. In autonomous systems, the past implies the future. In controllable systems, the past has no lasting implications about the far future.

Let $\Sigma = (T, W, \mathfrak{B})$ be a time invariant system with $T = \mathbb{R}$ or $\mathbb{Z}$. We will call it *autonomous* if there exists a map $f: \mathfrak{B}^- \to \mathfrak{B}^{0+}$ such that for all $w \in W^T$, there holds $\{w = w^- \wedge_{0^-} w^{0+} \in \mathfrak{B}\} \Leftrightarrow \{w^- \in \mathfrak{B}$ and $w^{0+} = f(w^-)\}$. It is said to be *controllable* if for all $w_1, w_2 \in \mathfrak{B}$ there exists a $t \in T$, $t \geq 0$, and a $w: T \cap [0, t) \to W$ such that $w_1 \wedge_{0^-} w \wedge_{t^-} \sigma^{-t} w_2 \in \mathfrak{B}$ (see Fig. 7). Note that in these definitions we could equivalently have demanded the existence of $f: \mathfrak{B}^{-0} \to \mathfrak{B}^{0+}$ for autonomous systems and any of the concatenations $w_1 \underset{0^-}{\wedge} w \underset{0^+}{\wedge} \sigma^{-1} w_2$, $w_1 \underset{0^+}{\wedge} w \underset{t^-}{\wedge} \sigma^{-t} w_2$, or $w_1 \underset{0^+}{\wedge} w \underset{t^-}{\wedge} \sigma^{-t} w_2$ for controllability.

The notion of controllability played an instrumental role in the advances in control and filtering of the early 1960s. Note, however, that in our point of

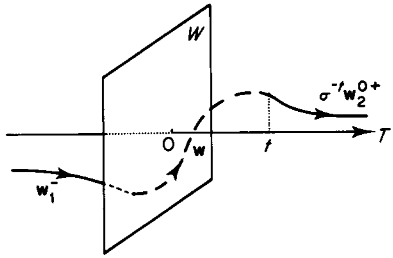


Fig. 7

view *controllability is a property of the external behaviour of a dynamical system*! In a controllable system we can, whatever be the past trajectory realized by 'nature', steer the system so that it will generate any desired future trajectory. This obviously is a natural formalization of the ability to steer the system in a desirable way: controllability.

Consider the difference equation of section 1.4.1. If this difference equation can be solved for the largest lag, i.e., if it is of the form

$$w(t+L)=f'(w(t+L-1),\ldots,w(t))$$

then the resulting dynamical system will be autonomous. Similarly, the continuous time system described by

$$\frac{d^n w}{dt^n}(t)=f'\left(\frac{d^{n-1}w}{dt^{n-1}}(t),\ldots,w(t)\right)$$

will be autonomous (assuming f' to be sufficiently smooth so that the differential equation possesses a unique solution for all initial data

$$\left(w(0),\frac{dw}{dt}(0),\ldots,\frac{d^{n-1}w}{dt^{n-1}}(0)\right).$$

The notions of linearity, symmetry, autonomy, controllability, etc., can be extended in an obvious way to state space systems, or to systems involving latent variables. In this context we will prove the following simple propositions which give insight in the concept of state and the notion of autonomous system, and in the relation of the classical notion of controllability to ours.

PROPOSITION 1.2 The state space system $\Sigma_s=(T,W,X,\mathfrak{B}_s)$ is autonomous if and only if there exists a map $\tilde{f}:X\to\mathfrak{B}_s^{0+}$ such that $\{(w,x)\in\mathfrak{B}_s\}\Rightarrow\{(w,x)^{0+}=\tilde{f}(x(0))\}$.

Proof The 'if' part is obvious. To show the 'only' if' part, consider $\mathfrak{B}_s(a):=\{(w,x)\in\mathfrak{B}_s \mid x(0)=a\}$. By the state axiom, $\mathfrak{B}_s^-(a)\Lambda_0\text{-}\mathfrak{B}_s^{0+}(a)=\mathfrak{B}_s(a)\subseteq\mathfrak{B}_s$. On the other hand, if $\mathfrak{B}_s$ is autonomous, there exists a map $f:\mathfrak{B}_s^-(a)\to\mathfrak{B}_s^{0+}(a)$ such that $\{(w,x)\in\mathfrak{B}_s(a)\}\Leftrightarrow\{(w,x)^{0+}=f((w,x)^-)\}$. This implies that $\mathfrak{B}_s^{0+}(a)$ consists at most of one element, which yields the result claimed. ■

Consider the (time invariant) dynamical system $\Sigma = (T, W, \mathfrak{B})$ with $T = \mathbb{Z}$ or $\mathbb{R}$. We will call Σ *trim* if for all $w \in W$ there exists a $\mathbf{w} \in \mathfrak{B}$ such that $\mathbf{w}(0) = w$. In a trim system all the external attributes can somehow occur. In an instantly specified system it is precisely the lack of trimness which expresses the laws of the system. This notion is easily generalized to systems with latent variables. As such we will speak about a system which is trim in the latent variables when the system $(T, W, P_a\mathfrak{B}_a)$ is trim. Since we view latent variables as auxiliary variables it is reasonable to assume that an internal behaviour is trim in the latent variables: if not, we can simply redefine the set A. When the latent variables are state variables, then we will speak about systems which are state trim.

The relation between our concept of controllability and the classical state controllability is as follows. Let $\Sigma_x = (T, X, \mathfrak{B}_x)$ be the state behaviour of a system. In other words, assume that it is Markovian. We will call Σ_x *point controllable* if, $\forall\, x_0, x_1 \in X$, there exists an $\mathbf{x} \in \mathfrak{B}_x$ and a $t \in T$, $t \geq 0$, such that $\mathbf{x}(0) = x_0$ and $\mathbf{x}(t) = x_1$. The concept of controllability as it is classically used in the control theory literature corresponds roughly speaking to point controllability of a state behaviour. We have the following result.

PROPOSITION 1.3 Let Σ_x be trim. Then it is controllable if and only if it is point controllable. Further, if Σ_x is point controllable, then Σ is controllable.

Proof Obvious. ■

1.5 Evolution laws

1.5.1 *Evolution laws*

Most models which one encounters in practice are in the form of differential or difference equations. It is well known that higher order differential or difference equations may be reduced to first order equations simply by redefining lagged variables or derivatives as new variables. We will consequently now treat only *first order* differential or difference equations. As we shall see, such models are automatically in state space form.

DEFINITION 1.4 A *discrete time evolution law* is defined as a quadruple

$$\Sigma_\partial = (T, W, X, \partial)$$

with $T \subseteq \mathbb{Z}$ the *time axis*; in the paper, $T = \mathbb{Z}$; $W =$ the *signal alphabet*; X the *state space*; and $\partial \subseteq X \times W \times X$ the *next state relation*.

The intuitive interpretation of ∂ is as follows: $(x_0, w, x_1) \in \partial$ signifies that if the system is in state x_0, then it can proceed to state x_1 while producing the external signal value w. Define the *behaviour induced* by ∂ as $\mathfrak{B}_\partial := \{(w,x):\mathbb{Z} \to W \times X \mid (x(t), w(t), x(t+1)) \in \partial$ for all $t \in \mathbb{Z}\}$. It is straightforward to verify that $\mathfrak{B}_\partial$ satisfies the axiom of state and that it is time invariant. From $\mathfrak{B}_\partial$ we obtain $\Sigma_s = (\mathbb{Z}, W, X, \mathfrak{B}_\partial)$, the *state space system induced* by Σ_∂, $\mathfrak{B} = P_w\mathfrak{B}_\partial$, the *external behaviour* and $\Sigma = (\mathbb{Z}, W, P_w\mathfrak{B}_\partial)$, the *dynamical system induced* by Σ_∂. We will denote this as:

$$\partial \Rightarrow \mathfrak{B}_\partial \Rightarrow \mathfrak{B} \quad \text{and} \quad \Sigma_\partial \Rightarrow \Sigma_s \Rightarrow \Sigma$$

The continuous time analogon of a next state relation is a first order differential relation: rather than telling where the state is allowed to go, we specify in what direction and with what velocity it can proceed.

DEFINITION 1.5 A *continuous time evolution law* is defined as a quadruple

$$\Sigma_\partial = (T, W, X, \partial)$$

with $T \subseteq \mathbb{R}$ an interval, the *time axis*; in this paper, $T = \mathbb{R}$; W the *signal alphabet*; X the *state space*, a differentiable manifold; and $\partial \subseteq TX \times W$ the *vector field relation* (TX denotes the tangent bundle of X).

For the purpose of the present paper it suffices to think of X as an open subset of $\mathbb{R}^n$ and identify TX with $X \times \mathbb{R}^n$.

Intuitively, $((x,v),w) \in \partial$ means that when the system is in state x, it will be able to evolve with velocity v while producing the external signal value w. Define the *behaviour induced* by ∂ as $\mathfrak{B}_\partial := \{(w,x):\mathbb{R} \to W \times X \mid x$ is absolutely continuous and $((x(t), \dot{x}(t)), w(t)) \in \partial$ for all $t \in \mathbb{R}$ where $\dot{x}(t)$ exists$\}$. Informally hence we can think of the behaviour of a discrete time evolution law as being defined as the solution set of a difference equation which is first order in x and zeroth order in w: $f(x(t), w(t), x(t+1) = 0(\partial := f^{-1}))$ while a continuous time evolution law can be thought of as being described by a differential equation which is first order in x and zeroth order in w: $f(x(t), \dot{x}(t), w(t)) = 0(\partial := f^{-1}(0))$. An example of a class of dynamical systems described by an evolution law are the systems governed by what are called *differential inclusions*: $\dot{x} \in f \circ (x)$; $w = h \circ (x)$ (or $w \in h \circ x$)), where f (and h) is a point to set map.

It is easily verified that $\mathfrak{B}_\partial$ satisfies the axiom of state. As in discrete time all this eventually leads to:

$$\partial \Rightarrow \mathfrak{B}_\partial \Rightarrow \mathfrak{B} \quad \text{and} \quad \Sigma_\partial \Rightarrow \Sigma_s \Rightarrow \Sigma$$

The convenience of specifying a behaviour by means of an evolution law may be explained as follows. If we define a system in terms of its behaviour then we basically give only a *rule*, a *specification*, a *law*, through which we can verify whether or not a particular time trajectory in W is or is not compatible with the system. An evolution law ∂ on the other hand, gives us a *grammar*, a *procedure*, an *algorithm* by means of which elements of $\mathfrak{B}$ can be generated.

Thus whether or not a pair (w, x) is compatible with the behaviour can be checked completely by means of the values at adjacent points, that is in terms of the local behaviour (with local to be understood as *local in time*—however, similar ideas are being pursued when also spatial variables are involved).

1.5.5 *The evolution law induced by a state space system*

The question which we will now discuss is a simple one: how to construct the evolution law which simulates a state space system. We will consider primarily the discrete time case.

Let $\Sigma_s = (\mathbb{Z}, W, X, \mathfrak{B}_s)$ be a (time invariant) discrete time state space system. Define the evolution law induced by Σ_s as $\Sigma_\partial := (\mathbb{Z}, W, X, \partial)$ with

$$\partial := \{(x_0, w, x_1) \in X \times W \times X \mid \exists (w, x) \in \mathfrak{B}_s \text{ such that } x(0) = x_0,\ x(1) = x_1, \text{ and } w(0) = w\}.$$

Of course, as we have seen in section 1.5.1 ∂ will induce a state space system. Denote the behaviour of this state space system by $\bar{\mathfrak{B}}_s$. Clearly $\mathfrak{B}_s \subseteq \bar{\mathfrak{B}}_s$. An example of a situation where strict inequality holds is $l_2(\mathbb{Z}; \mathbb{R}^q)$ [more precisely, the state space system $(\mathbb{Z}, \mathbb{R}^q, 0, l_2(\mathbb{Z}; \mathbb{R}^q))$].In this case $\bar{\mathfrak{B}}_s$ equals all of $(\mathbb{R}^q)^{\mathbb{Z}}$ which includes $l_2(\mathbb{Z}; \mathbb{R}^q)$ as a strict subset. The question thus arises: *when is* $\mathfrak{B}_s = \bar{\mathfrak{B}}_s$?

Let $\Sigma = (T, W, \mathfrak{B})$ be a dynamical system. The *completion* of its behaviour is defined by

$$\mathfrak{B}^{\text{completion}} := \{w : T \to W \mid w|_{[t_0, t_1]} \in \mathfrak{B}|_{[t_0, t_1]} \text{ for all } -\infty < t_0 \leq t_1 < \infty\}.$$

It is easily seen that $\mathfrak{B}^{\text{completion}}$ is the smallest subset of W^{T} which is complete and contains $\mathfrak{B}$. It is clear that $\mathfrak{B}^{\text{completion}}$ will be time invariant and/or linear if $\mathfrak{B}$ is. We can now state and prove the following result.

THEOREM 1.1 Let $\Sigma_s = (\mathbb{Z}, W, X, \mathfrak{B}_s)$ *be a state space system and* $\bar{\mathfrak{B}}_s$ *the behaviour of the evolution law induced by it. Then* $\bar{\mathfrak{B}}_s = \mathfrak{B}_s^{\text{completion}}$. *Hence* $\{\mathfrak{B}_s = \bar{\mathfrak{B}}_s\} \Leftrightarrow \{\Sigma_s$ *is complete*$\}$. *In other words, a state behaviour is faithfully represented by an evolution law if and only if it is complete.*

Proof Since $\mathfrak{B}_s$ is Markovian, $\mathfrak{B}_s^{\text{completion}}$ has memory span 1. By the results of section 1.4.1, $\mathfrak{B}_s^{\text{completion}}$ can be described by behavioural equations with first order lag. Let $f(x(t), w(t), x(t+1), w(t+1)) = 0$, with $f^{-1}(0) = \mathfrak{B}_s^{\text{comp}}|_{[0,1]} = \mathfrak{B}_s|_{[0,1]}$, be this equation. However, it follows from the axiom of state that $\{(x(t), w(t), x(t+1)) \in \partial$ and $(x(t+1), w(t+1), x(t+2)) \in \partial\} \Rightarrow \{((w(t), x(t)), (w(t+1), x(t+1)) \in \mathfrak{B}_s|_{[0,1]}\}$.

Hence $\partial = \{x_0, w_0, x_1 \mid \exists w_1 \text{ such that } f(x_0, w_0, x_1, w_1) = 0\}$. The result follows. ■

Let $\Sigma_s = (\mathbb{R}, W, X, \mathfrak{B}_s)$ be a (time invariant) continuous time dynamical system in state space form with X a differentiable manifold, having the property that $\{(w,x)\in\mathfrak{B}_s\}\Rightarrow\{x \text{ is absolutely continuous}\}$. Define the evolution law induced by Σ_s as $\partial:=\{((x,v),w)\in TX\times W\,|\,\exists(w,x)\in\mathfrak{B}_s$ such that $(x(0),\dot{x}(0))=(x,v)$ and $w(0)=w\}$. With $\bar{\mathfrak{B}}_s$ defined analogously as in the discrete time case, there still holds $\mathfrak{B}_s\subseteq\mathfrak{B}_s^{\text{completion}}\subseteq\bar{\mathfrak{B}}_s$. However, in order to achieve equality of the later two, other conditions in addition to completeness must be satisfied. These issues, related to smoothness, will not be pursued here.

1.5.3 *The evolution law of a deterministic system*

One more definition: We will call the state space system $\Sigma_s=(T,W,X,\mathfrak{B}_s)$*(state) deterministic* if $\{(w_1,x_1),\ (w_2,x_2)\in\mathfrak{B}_s,\ x_1(0)=x_2(0),\ t\in T$, and $w_1|_{[0,t)}=w_2|_{[0,t)}\}\Rightarrow\{x_1(t)=x_2(t)\}$. In order words determinism means that state trajectories can only bifurcate as a consequence of a bifurcation of the external trajectory.

Observe that if a system described by a discrete time evolution law is deterministic, then ∂ equals the graph of a partial map $\gamma:X\times W\to X$, signifying that $\{(a,w,b)\in\partial\}\Leftrightarrow\{(a,w)\in Do(\gamma)$ and $b=\gamma(a,w)\}$. Now introduce two maps $f:X\times W\to X$ and $c:X\times W\to\mathbb{R}$ such that $c(x,w)=0$ defines the domain of γ and f corresponds to the action of γ on its domain. This shows that a discrete time evolution law of a deterministic state space system is described by a *next state map* $f:X\times W\to X$ and a *constraint equation* $c:X\times W\to\mathbb{R}$ such that its behaviour will be specified by

$$\sigma x=f\circ(x,w);\quad c\circ(x,w)=0.$$

In continuous time systems this will lead to the equations

$$\dot{x}=f\circ(x,w);\quad c\circ(x,w)=0.$$

These expressions yield a convenient way of thinking about state space systems. They represent deterministic complete state systems. The first equation tells us how a realization of the external signal variables will cause the state to evolve, while the second equation tells us which external signal variables w can actually occur when the system is in state x.

1.5.4 *Flows*

Examples of dynamical systems described by an evolution law are the 'classical' dynamical systems in which the state evolves in an autonomous way. A *discrete time flow* (X,f) is defined by a *state space* X and a *next state map* $f:X\to X$. A *continuous time flow* (X,f) is defined by a *state space* X, a differentiable manifold, and a *vector-field* $f:X\to TX$ on it. Flows define special cases of evolution laws with $W=X$ and

$\partial=\{(x_0,w,x_1)\,|\,w=x_0$ and $x_1=f(x_0)\}$ (discrete time)
behavioural equation: $\sigma x=f\circ(x)$

and

$$\partial = \{((x,v),w) \mid w = x \text{ and } (x,v) = f(x)\} \qquad \text{(continuous time)}$$
behavioural equation: $\dot{x} = f\circ(x)$

where we have identified, *somewhat artificially*, the external signal with the state. We also need to assume that for any initial condition, the differential equation $\dot{x} = f\circ(x)$; $x(0) = x_0$, has a unique solution. Flows clearly define autonomous systems (viewed as a property of the behaviour $\mathfrak{B}_s$). In fact, they are Markovian and hence state space systems.

It follows immediately from the first proposition in section 1.4.5 that an autonomous state space system is always deterministic. Its evolution law is expressed by

$$\sigma x = f\circ(x); \quad w = r\circ(x) \qquad \text{(discrete time)}$$
$$\dot{x} = f\circ(x); \quad w = r\circ(x) \qquad \text{(continuous time)}$$

We can hence think of an evolution law for an autonomous state space system as a flow together with a read-out map $r: X \to W$ (an '*observed flow*', if you like).

Flows on manifolds have often been proposed as the basis for dynamical models in physics. Indeed, Hamiltonian mechanics and the Schrödinger equation of quantum mechanics define, as we shall see, flows on manifolds (often with, implicitly, a non-trivial read-out map). This may make it seem appealing to try to develop flows as a basis for dynamics, at least for mechanics. This is, in fact, what has been done. However, in our opinion, this point of view suffers from two serious drawbacks.

First, because they define autonomous systems, flows consider the system in isolation from its environment. Not only is this very limiting as far as applications are concerned since often it is precisely the action and reaction of systems with their environment which is of central importance. In control theory and computer science, this is evident. However, also in physics there are many situations of this nature. Moreover, this assumption of isolating a system from its environment implicitly forces us to make a model of the reaction of the environment on the system, and so, whenever we model a system as a flow, we find ourselves forced in the unwanted and undesirable situation of having to also model the environment!

Second, models which start with flows on manifolds consider the state space as given, whereas we consider the external behaviour as essential and the state as a convenient *mathematical* object which is to be constructed on the basis of the dynamical equations which describe the external behaviour. The state of a system is not a phyical property of a real life system, it is a property of a model. Modelling slightly more accurately can and will have dramatic effects on the nature of the state space. If one models the planets in the solar system as point masses then one obtains a 22-dimensional state space. If on the other hand one considers one of the planets as a slightly elastic sphere, then the state space will already become infinite-dimensional. The logic of modelling by means of flows on manifolds reads: *first construct the state space* X, then construct the dynamical equations, *the vector field* f. However, since it is the dynamical equations which should tell us what the state space X is, this logic is circular. The logic of modelling by means of the behaviour as in Definition 1.1 reads: *first select what you want to model: choose* W, *then construct* $\mathfrak{B}$, *then, if required or desired, construct* X.

1.6 More examples

1.6.1 *A word example*

Let us illustrate the discussion about isolating a system from its environment by means of a verbal *Gedankenbeispiel* illustrating what we mean by a '*system*' by its '*environment*'.

We will consider the *flight of a bird*. If we consider the position of the bird as the primary variable of interest then, in order to describe the evolution of this position, we will have to introduce (at least) the motion of its wings and the conditions of the atmosphere around the bird (for example the wind speed and direction) as additional variables. The resulting compatibility relation among these variables will describe the flight of the bird. As a model, this is an appropriate point to stop. It explains the position of the bird within its environment consisting of the motion of its wings and the wind characteristics. This model obviously involves unexplained variables: the motion of the wings and the wind characteristics.

In a more ambitious modelling effort, however, we may want to include a model for the atmospheric conditions (for example by assuming that the wind speed and direction are constant or a given function of the height). This will lead to a compatibility relation involving as variables the position of the bird and motion of its wings. As a model for the flight of the bird this is, again, an appropriate point to stop. It explains the relation between the position of the bird and its environment consisting of the movement of its wings.

We can be even more ambitious modellers and try to explain also the motion of the wings. At this point no physical theory will tell us how to proceed: invariably this step will bring us outside the descriptive realm of physics into the prescriptive sciences sometimes called '*Cybernetics*' or the '*Sciences of the Artificial*'. Indeed, somehow we will have to explain why the bird moves its wings the way it does. One could do this by postulating a periodic motion for the wings. Undoubtedly, much will be learned by studying the resulting system of equations under this assumption, however naive. A more sophisticated approach would be to deduce the movement of the wings by making the bird into a *purposeful system*: say, if the bird is a predator, reaching its prey, a rodent, in minimum time. The resulting model will be a compatibility relation connecting the position of the bird with that of the rodent. This is, once more, an appropriate point to stop. It explains the position of the bird in its environment consisting of the position of the rodent.

Let us play this game one more set. One may also want to model the position of the rodent. This problem is similar to that of the bird. Its position will be a function of the motion of its legs and of the terrain. We could model the terrain and we could also make the rodent into a purposeful system: say that the motion of its legs can be explained by the maximization of the distance from its predator, the bird. This will yield a model for the position of the rodent against its environment consisting of the position of the bird. In total we would now have obtained two behavioural compatibility relations involving the positions of the bird and of the rodent. Together they are likely to give us a closed system of equations which determine the position of the bird as a function of the initial conditions.

What is the point of this example? Primarily we wanted to demonstrate what it means 'isolating' a system from its environment but considering it 'in interaction' with it. Invariably this will involve leaving some variables unexplained: these will come from the outside and are, in principle, arbitrary. Such unexplained time functions are an almost unavoidable part of mathematical models of dynamical systems. Our example also shows that rather simple situations, as the one described, will already involve an interconnection of a number of physical and cybernetic subsystems.

This example involves a living system, but that is not important. Cars, bicycles, windmills, economies are other examples which can reasonably only be described by allowing unexplained external influences.

1.6.2 *Kepler's laws*

According to Kepler, the motion of planets in the solar system obey the following three laws:

K.1: They move in elliptical orbits with the sun in one of the foci;
K.2: The radius vector from the sun to the planet sweeps out equal areas in equal times;
K.3: The square of the period of revolution is proportional to the third power of the major axis of the ellipse.

This defines a dynamical system with (disregarding biblical considerations) $T = \mathbb{R}$, $W = \mathbb{R}^3$, and $\mathfrak{B}$ the family of all orbits satisfying K.1, K.2 and K.3. This system is time invariant, nonlinear, autonomous, and locally specified (the trajectories are analytic), hence with local memory; it is not Markovian (consequently finding a convenient state representation in principle presents a problem). This system is, moreover, time-reversible and has the subgroup of $\mathscr{G}l(3)$ consisting of $\{L \in \mathscr{G}l(3) \mid |\det L| = 1\}$ as a symmetry in the sense of 1.2.3(ii).

1.6.3 *Hamiltonian mechanics*

We will not describe here the elegant and natural setting of Hamiltonian mechanics in terms of symplectic geometry, but limit our attention to situations in which the configuration space Q is an open subset of $\mathbb{R}^m$. According to the postulates of Hamiltonian mechanics, the motion of a mechanical system may be described by a single function $H: P \times Q \to \mathbb{R}$, with $P = \mathbb{R}^m$, the momenta space. This function H is the *Hamiltonian* and it determines the laws of motion via the canonical equations

$$\dot{q} = \frac{\partial H}{\partial p}(p, q) \tag{H1}$$

$$\dot{p} = -\frac{\partial H}{\partial q}(p, q) \tag{H2}$$

Assume existence and uniqueness of a solution of this set of differential equations for any initial condition $\boldsymbol{p}(0)=p_0\in\mathbb{R}^n$ and $\boldsymbol{q}(0)=q_0\in Q$. The equations ($H$) obviously define a flow on the manifold $P\times Q$. Formally $T=\mathbb{R}$(or $\mathbb{R}_+$)

$$X=P\times Q \quad\text{and}\quad f=\left(-\frac{\partial H}{\partial q},\frac{\partial H}{\partial p}\right).$$

If, however, we view these equations as a convenient way of describing the evolution of the position q, with the momentum p considered as an auxiliary variable, then we arrive at a system with latent variables, with $T=\mathbb{R}$, $W=Q$, $A=P$, and $\mathfrak{B}_a=\{(\boldsymbol{q},\boldsymbol{p})|(H)$ is satisfied$\}$. This system with auxiliary variables has (p,q) as its state. The external behaviour is $\mathfrak{B}=\{q|\exists\boldsymbol{p}$ such that (H) is satisfied$\}$. This system is time invariant. It is time reversible if $H(p,q)=H(-p,q)$. It is likely that this system is autonomous (although we know of no formal proof of this, unless H is a quadratic form, in which case the system is linear).

The above definition of $\mathfrak{B}$ implies that we are primarily interested in the position q. If we are also interested in the velocity then we can simply add the equation

$$\boldsymbol{v}=\dot{\boldsymbol{q}} \tag{H3}$$

yielding the external behaviour $\mathfrak{B}=\{(\boldsymbol{q},\boldsymbol{v}):\mathbb{R}\to Q\times\mathbb{R}^n|\exists\boldsymbol{p}:\mathbb{R}\to\mathbb{R}^n$ such that (H1), (H2) and (H3) hold$\}$. If $H(p,q)=H(-p,q)$ then this system is dynamically time reversible in the sense of section 1.2.3(iv) with $J(q,v)=(q,-v)$.

1.6.4 *Quantum mechanics*

Quantum mechanics warns us not to speak lightly about the position of a particle as a physical reality but instead to ponder about the 'probability' of finding a particle in a certain region of space $\mathbb{R}^3$. Thus we will obtain a dynamical system with time axis $T=\mathbb{R}$ and signal alphabet $P:=\{p:\mathbb{R}^3\to\mathbb{R}|p\geq 0$ and $\int_{\mathbb{R}^3}p(z)dz=1\}$: this is the collection of all probability measures (which for simplicity we have taken to be absolutely continuous w.r.t. Lebesgue measure) on $\mathbb{R}^3$. In order to specify the behaviour it has proven to be convenient to introduce the *wave function* $\psi:\mathbb{R}^3\to\mathbb{C}$ as an latent variable. Thus define the space of latent variables $\Psi:=\mathscr{L}_2(\mathbb{R}^3;\mathbb{C})$. The internal behaviour $\mathfrak{B}_a\subseteq(P\times\Psi)^{\mathbb{R}}$ is defined by two relations. The first one determines p as a function of ψ and the second one, the *Schrödinger equation*, tells us how ψ evolves in time. Let $\boldsymbol{\psi}:\mathbb{R}^3\times\mathbb{R}\to\mathbb{C}$ be the time trajectory of the wave function. Hence $\boldsymbol{\psi}(z_1,z_2,z_3;t)$ denotes the value of the wave function at the position $(z_1,z_2,z_3)\in\mathbb{R}^3$ at time $t\in\mathbb{R}$. Similarly, let $\boldsymbol{p}:\mathbb{R}^3\times\mathbb{R}\to\mathbb{R}_+$ denote the time trajectory of the probability density function. The wave function generates the probability density by

$$\boldsymbol{p}(z_1,z_2,z_3;t):=\frac{|\boldsymbol{\psi}(z_1,z_2,z_3;\ t)|^2}{\displaystyle\int_{\mathbb{R}^3}|\boldsymbol{\psi}(x_1,z_2,z_3;t)|^2dz_1dz_2dz_3}. \tag{QMI}$$

The evolution of the wave function is governed by Schrödinger's equation:

$$-i\frac{h}{2\pi}\frac{\partial\boldsymbol{\psi}}{\partial t}=H(\boldsymbol{\psi}) \tag{QM2}$$

where the *Hamiltonian* H is a linear, in general unbounded, operator on $\mathscr{L}_2(\mathbb{R}^3;\mathbb{C})$, and h is Planck's constant. The Hamiltonian is specified by the potential and the geometry and should be considered as fixed for a given system. This yield the extended behaviour

$$\mathfrak{B}_a := \{(p, \psi): \mathbb{R} \to P \times \Psi \mid (\text{QM}) \text{ is satisfied}\}$$

which we view as a convenient way of specifying the external behaviour

$$\mathfrak{B}: \{p: \mathbb{R} \to P \mid \exists \psi: \mathbb{R} \to \Psi \text{ such that (QM) is satisfied}\}$$

Clearly this system $(\mathbb{R}, W, \mathfrak{B})$ is time invariant and, most likely, it is also autonomous (although we know of no formal proof of this). The system $(\mathbb{R}, P, \Psi, \mathfrak{B}_a)$ is an autonomous state space system. If we restrict our attention to the wave function alone, i.e., if we consider the dynamical system $(\mathbb{R}, \Psi, P_\psi \mathfrak{B}_a)$, then we obtain a linear flow.

The point of view taken here, in which ψ is a latent variable aimed at modelling p, is a very logical one indeed. The truly surprising fact however is that the (very *nonlinear*) behaviour $\mathfrak{B}$ can be represented by means of a *linear* flow (QM2), the Schrödinger equation, together with the *memoryless* map, the static behavioural equation (QM1). Note, however, that the point of view that ψ is introduced in order to model p, however logical, does not do justice to the historical development in which ψ had been studied long before the probability interpretation of $|\psi|^2$ was suggested.

Note finally that our approach discusses probability in a purely deterministic tone—stochastic generalizations of the concepts developed in this paper are another story altogether.

1.6.5 *Discrete event systems*

Discrete event systems describe situations in which the occurrence of previous events enables or blocks the occurrence of subsequent events. Think for instance of natural languages, computer codes, manufacturing systems, committee meetings (even though in first instance we think of discrete event systems as describing an orderly sequence of events), etc. In our thinking, following Definition 1.1, we will speak of a *discrete event system* simply as a dynamical system $(T, W, \mathfrak{B})$ with $T = \mathbb{Z}$ and W a *finite* set, and, if the system is in state space form (or is defined in terms of latent variables), with X (or the set of latent variables) finite. We now want to show how one can view the concept of a formal language in this setting.

Consider a non-empty *finite* set A, called the *alphabet*, whose elements are called *symbols*. A finite (possibly empty) string of symbols is called a *word*. Let A^* be the set of all words consisting of symbols from the alphabet A. Erudite individuals refer to A^* as the *free monoid generated by A*. A *formal language*, $\mathscr{L}$, is simply a subset of A^*. We think of elements of $\mathscr{L}$ as legal words: those words compatible with the rules, the grammar, the laws, governing the language. The above nomenclature is clearly borrowed from natural languages. We could also call A the *event set*, elements of A *elementary events*, elements of A^* *event strings* (or *traces*), and think of $\mathscr{L}$ as the collection of all feasible (finite) sequences of events; $(A, \mathscr{L})$ is sometimes called a *trace structure*. A formal language basically defines a dynamical system in the sense of Definition 1.1, with $\mathscr{L}$ corresponding to the behaviour $\mathfrak{B}$. However, in order to make this correspondence hard we need to apply some minor cosmetics in order to make sure that all words are equally long and that no new words are introduced in the process. In order to do this, add a new symbol, the *blank*, $\square$, to A, define $W := A \cup \{\square\}$ and

$$\mathfrak{B} := \{w: \mathbb{Z} \to W \mid \exists t_{-1}, t_1 \in \mathbb{Z},\ t_{-1} \leq t_1, \text{ such that } w|_{[t_{-1}, t_1)} \in \mathscr{L} \\ \text{and } w(t) = \square \text{ for } t < t_{-1} \text{ and } t \geq t_1\}.$$

Clearly $(\mathbb{Z}, W, \mathfrak{B})$ defines a time invariant dynamical system in the sense of Definition 1.1, with $\mathfrak{B}$ deduced from $\mathscr{L}$ in a simple one-to-one way. All we have done is add an infinite number of blanks to the front and back of every word. This illustrates that our notion of a discrete event system is a simple and natural generalization of the notion of a formal language (for discrete event systems our definition adds the possibility of infinite words).

Common procedures for generating formal languages are by means of grammars (which vaguely corresponds to describing systems by means of latent variables) and by automata which basically corresponds to our evolution laws). We will explain automata in the next section. Note, however, that in discrete event systems one usually should interpret the time index t as *logic time* (meaning that it merely parametrizes the sequencing of the events) in contrast to the usual interpretation in physics and economics where t denotes *clock time.*

1.6.6 *Automata*

We will now describe automata, discrete event systems in state space form. An *automaton* is a quintuple (S, A, E, I, F) and S a finite set called the *state space*; A a finite set called the *alphabet*, its elements are called *symbols* (or *elementary external events*); E the *state transition rule*: E is a subset of $S \times A \times S$ and its elements are called *edges* (or *elementary internal events*); $I \subseteq S$ is the set of *initial states*; $F \subseteq S$ is the set of *terminal states.* A sequence $(s_0, a_0, s_1, a_1, \ldots, s_{n-1}, a_{n-1}, s_n)$, with $(s_i, a_i, s_{i+1}) \in E$ for $i+1 \in \mathbf{n}$ is called a *path*; it is called a *successful path* if in addition $s_0 \in I$ and $s_n \in F$.

Automata are usually represented by means of directed graphs with the states as nodes, the edges as branches labelled with the corresponding symbol, initial states as nodes with an arrow pointing towards it, and terminal states as nodes with an arrow pointing away from it. This is illustrated by means of a binary adder in Fig. 8. This automaton achieves the addition of two binary numbers: these are coded in the first two symbols next to the arrows. The sum is coded in the last symbol. The state is the memory acquired in the sequential addition.

However, in order to make this correspondence hard, add again the *blank*, $\square$, to A, yielding $W := A \cup \{\square\}$, add two states, a *source state*, $\circ\!\!\rightarrow$, and a *sink state*, $\rightarrow\!\!\circ$, to S, yielding $X := SU\{\circ\!\!\rightarrow, \rightarrow\!\!\circ\}$, and define the evolution law $\partial \subseteq X \times W \times X$ as $\partial := E \cup \{\circ\!\!\rightarrow, \square, I \cup \{\circ\!\!\rightarrow\})\} \cup \{(F \cup \{\rightarrow\!\!\circ\}, \square, \rightarrow\!\!\circ\}$. Of course, our modification of the original automaton is an automaton in its own right. For the binary adder, this modification is shown in Fig. 9.

Now define the behaviour of the (modified) automaton as $\mathfrak{B}_s = \{(w, x): \mathbb{Z} \to W \times X \mid (x(t), w(t), x(t+1)) \in \partial$ for all $t \in \mathbb{Z}$ and $\exists t_{-1}, t_1, t_{-1} \leq t_1$, such that $x(t) = \circ\!\!\rightarrow$ for $t < t_{-1}$ and $x(t) = \rightarrow\!\!\circ$ for $t \geq t_1\}$. It is easily seen that $\mathfrak{B}_s$ basically consists of the successful paths with

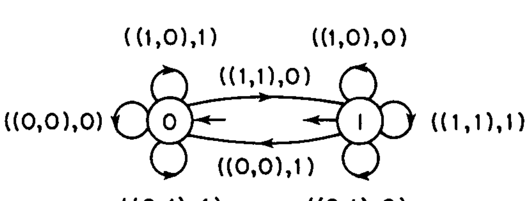


Fig. 8

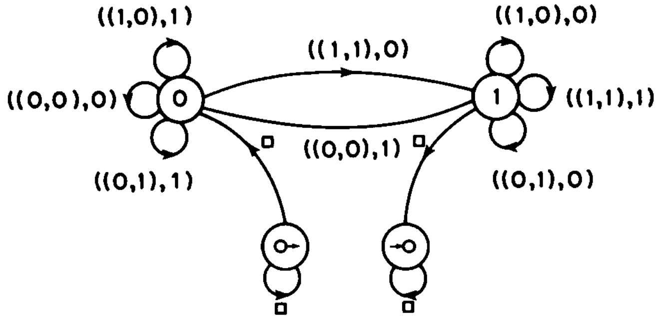


Fig. 9

an infinite string of (○→, □) pairs added to the front and an infinite string of (□, →○) pairs added to the back of every successful path. This shows that there is a one-to-one correspondence between $\mathfrak{B}_s$ and the collection of successful paths. A word $a_1 a_2 \cdots a_n$ with $a_i \in A$, $i \in \boldsymbol{n}$, is said to be recognized by the automation if there exists a sequence $s_0, s_1, \ldots, s_n$ with $s_i \in S$, $i = 0, 1, \ldots, n$, such that $(s_0, a_0, s_1, a_1, \ldots, s_{n-1}, a_{n-1}, s_n)$ is a successful path. The collection of all words recognized by an automaton is a formal language. Thus the question arises of how to construct an automaton which recognizes a given language. This problem is a version of the state space representation problem which will be discussed in section 2. The difficulty in the context of formal languages is to clarify the conditions on $\mathscr{L} \subseteq A^*$ which guarantee that it has the internal structure allowing it to be represented by means of a *finite* automaton (an automaton having only a finite number of states).

Formal languages and automata in their traditional setting are examples of non-complete systems: the finiteness of the words in a formal language (equivalently, the blanks at the beginning and at the end of every word of the modified language) and the initial and terminal state constraints in automata (equivalently, the presence of the source state and the sink state in the modified automaton) are typical conditions preventing completeness. We will now make this precise.

Let (S, A, E, I, F) be an automaton. We will say that all its states are *attainable* if for all $s \in S$ there exists a path, $(s_0, a_0, s_1, \ldots, s_{n-1}, a_{n-1}, s_n)$ with $s_0 \in I$ and $s_n = s$. We will say that it contains no *dead-end states* if instead we can take $s_0 = s$ and $s_n \in F$. An automaton in which all states are attainable and which contains no dead-end states is state trim. We will call a path $(s_0, a_0, s_1, \ldots, s_{n-1}, a_{n-1}, s_n)$ a *cycle* if $s_0 = s_n$. It is easy to see that if an automaton is trim then it is complete (i.e. $\mathfrak{B}_s = \mathfrak{B}_s^{\text{comp}}$) if and only if it contains no cycles. Another way of saying this is that a trim automaton is complete if and only if the modified automaton contains no livelocks (a *livelock* is a cycle which does not pass through a terminal state). This condition can be translated in terms of the formal language recognized by the automaton. Indeed, a *trim automaton will be complete if and only if the language which it recognizes is finite* (i.e., if and only if $\mathscr{L}$ contains only a finite number of words). This shows that completeness of automata is very exceptional and that the initial and terminal states are, as well as the evolution law, essential for the description of the functioning of an automaton.

The set of complete discrete event systems becomes much larger if we do not insist on having only words of finite length in $\mathscr{L}$. This classical restriction of finiteness of all words is reasonable in many applications in computer science, but there are also many situations of discrete event systems (as counters, traffic lights, and other discrete event systems in continuous operation) where incorporation of the finiteness of all words in the model is all but natural.

1.6.7 *Recapitulation*

In this section we have introduced the model classes of the dynamical systems which will be considered in this paper. The basic object of the study is the *behaviour* of a dynamical system. This is a succinct way of formalizing a model, of formulating the laws which govern a system, of specifying which trajectories can and cannot occur. However, models which we write down from first principles will invariably involve *latent variables.* Such variables are also important in theoretical considerations. A prime example are *state space systems* in which the latent variable, the state, splits the past and future behaviour of the signal and the state jointly. This property is called the *axiom of state.* State models are often specified in terms of an *evolution law*, in which the dynamical laws governing a system are purely local in time. In discrete time, evolution laws take the form of difference equations which are first order in the state and zeroth order in the signal variable, and in continuous time, evolution laws take the form of differential equations which similarly are first order in the state and zeroth order in the signal variable. The relation between these different specifications of dynamical systems is schematically shown below:

$$\text{Evolution law} \Rightarrow \text{State models} \Rightarrow \text{Splitting variables} \Rightarrow \text{Latent variables} \Rightarrow \text{External behaviour}$$

In section 2 we will study the reversion of some of these arrows.

1.6.8 *Sources*

Definitions 1.1, 1.2, 1.3 and 1.4 form a plateau in a struggle to make suitable, general definitions for dynamical systems. The attempts coming from physics/mechanics/differential equations [1], [2] usually arrive at a version of flows on manifold discussed in section 1.5.4. The fact that such models ignore the interaction of a system with its environment severely limits their scope and applicability—even in mechanics. The attempts coming from control theory [3], [4], [5] invariably arrive at input/output maps or input/output relations. The input/output structure implies more structure than is, or needs to be, present in many dynamical systems. There also have been attempts coming from General Systems Theory [6], [7] and very sophisticated dynamical structures (formal languages, automata, etc.) have been studied in computer science [8]. The notion of state is basic in physics and is (almost trivially) incorporated in the definition of a flow. The study of the state together with the interactions with the environment is, in an input/output setting, one of the main contributions of modern control theory as expressed for example in the work of Kalman [9], Bellman [10], and Pontryagin [11]. The basic framework presented in this section (in particular Definitions 1.1 and 1.3) were first proposed in [12] and further developed in [13], [14]. The explicit introduction of latent variables in Definition 1.2 is an important refinement of our earlier work. Latent variables also appear, in a disguised form, in the work of

Rosenbrock [15] and are also used in computer science and mathematical linguistics in the context of production rules and grammars.

2 MODELS FROM OTHER MODELS—EXTRACTING THE MEMORY STRUCTURE OF A DYNAMICAL SYSTEM

In this section we will discuss methods for writing models for dynamical systems, which are given in terms of their behaviour, in state space form. This problem, called the *realization problem*, is conceptually one of the richest and one of the most researched problems in the mathematical system theory literature. The theory developed in this section is purely set theoretic in nature. In the fourth section we will discuss systems with more structure.

2.1 Observability

2.1.1 *Observability*

We will first introduce another important concept: that of observability. In the classical theory, observability is a property of a state space system. For us, it will be a property of the external behaviour. We will consider systems $\Sigma=(T, W, \mathfrak{B})$ defined on a product set $W=W_1\times W_2$. As usual $P_{w_1}: W_1\times W_2\to W_1$, and $P_{w_2}: W_1\times W_2\to W_2$ will denote the projections $P_{w_1}(w_1, w_2):=w_1$ and $P_{w_2}(w_1, w_2):=w_2$. As before, we will assume that P_{w_1} and P_{w_2} are also defined on $(W_1\times W_2)^T$. For simplicity we will use the notation $\mathfrak{B}_1:=P_{w_1}\mathfrak{B}$ and $\mathfrak{B}_2:P_{w_2}\mathfrak{B}$. Of course, $\mathfrak{B}\subseteq\mathfrak{B}_1\times\mathfrak{B}_2$ and the fact that $\mathfrak{B}$ is a strict subset of $\mathfrak{B}_1\times\mathfrak{B}_2$ specifies the connection which the laws of the dynamical system impose on the signals $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$. Also, note that when we consider a system with auxiliary variables $\Sigma_a=(T, W, A, \mathfrak{B}_a)$(or a state system) we can view it as the system $(T, W\times A, \mathfrak{B}_a)$ defined on the product set $W\times A$. We will frequently use this implicitly, for example when discussing observability of a state system. Let $\Sigma=(T, W_1\times W_2, \mathfrak{B})$ be a (time invariant) dynamical system. Hence each element of $\mathfrak{B}$ consists of a pair of time functions $(\boldsymbol{w}_1, \boldsymbol{w}_2)$, with $\boldsymbol{w}_1: T\to W_1$ and $\boldsymbol{w}_2: T\to W_2$. We will call w_2 *observable from* w_1 if there exists a map $F:\mathfrak{B}_1\to\mathfrak{B}_2$ such that $\{(\boldsymbol{w}_1, \boldsymbol{w}_2)\in\mathfrak{B}\}\Leftrightarrow\{\boldsymbol{w}_2=F\boldsymbol{w}_1\}$. If, moreover, this map F is non-anticipating then we will call w_2 *observable from the past* of w_1. If F is strictly non-anticipating then we will call w_2 *observable from the strict past* of w_1. If, finally, F is purely anticipating (that is, if $\{\boldsymbol{w}'_1, \boldsymbol{w}''_1\in\mathfrak{B}_1, \boldsymbol{w}'_1(t')=\boldsymbol{w}''_1(t')$ for $t'\geq t\}\Rightarrow\{(F\boldsymbol{w}'_1)(t')=(F\boldsymbol{w}''_1)(t')$ for $t'\geq t\}$) then will call w_2 *observable from the future* of w_1.

We prefer to use a slightly different nomenclature when applying these observability concepts to state space systems $\Sigma_s=(T, W, X, \mathfrak{B}_s)$(viewed as the system $(T, W\times X, \mathfrak{B}_s)$, with w playing the role of the observed variable: w_1 in the above definitions, and with x playing the role of the deduced variable: w_2 in the above definitions). Thus, if there exists a map F from $\mathfrak{B}:=P_w\mathfrak{B}_s$ to X such that $\{(\boldsymbol{w}, \boldsymbol{x})\in\mathfrak{B}_s\}\Rightarrow\{\boldsymbol{x}(0)=F\boldsymbol{w}\}$, then we will call the state *externally induced.*

Further, if there exists a map F^- *from* $\mathfrak{B}^- := (P_w\mathfrak{B}_s)^-$ to X such that $\{(w,x)\in\mathfrak{B}_s\} \Rightarrow \{x(0) = F^- w^-\}$, then we will call the state *past induced*. If, instead, there exists a map F^+ from $\mathfrak{B}^{0+} := (P_w\mathfrak{B}_s)^{0+}$ to X such that $\{(w,x)\in\mathfrak{B}_s\} \Rightarrow \{x(0) = F^+ w^{0+}\}$, then we will call the *state future induced*. Note that if the state is past induced, then the state system is deterministic. The concept of observability as it is classically used in the control theory literature corresponds roughly speaking to what we call a state space system with an externally induced state.

2.2 Construction of state representations

2.2.1. *The trivial realization*

We will now start studying the problem of constructing a state representation for a given system. We begin by formally defining the problem once more. Assume that $\Sigma = (T, W, \mathfrak{B})$ is a dynamical system (time invariant, of course, with $T = \mathbb{R}$ or $\mathbb{Z}$). The problem is to find a state space system $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ such that its external behaviour, defined as $\{w \mid \exists x \text{ such that } (w,x)\in\mathfrak{B}_s\}$, equals $\mathfrak{B}$. If this is the case, then we will call Σ_s a *state representation* or a *state realization*—we will use both terms as synonymous—of Σ or $\mathfrak{B}$.

The basic question, as we shall see, is the invention, the discovery, the construction, on the basis of $\mathfrak{B}$, of the state space X. However, as it now stands, the realization problem is trivial to solve. Simply take $X = \mathfrak{B}$ and define $\mathfrak{B}_s : \{(w,x): T \to W \times X \mid w\in\mathfrak{B} \text{ and } x(t) = \sigma^t w\}$. Clearly this defines a time invariant state representation of $\mathfrak{B}$. We will call it the *trivial realization*. This realization is very inefficient. For example, the state space X will be finite or finite-dimensional only in very exceptional circumstances and the trivial realization does not unfold the fine structure in the dynamics of $\mathfrak{B}$. Indeed, in trying to split the past from the future in the state behaviour, we have decided in the trivial realization to store the whole trajectory w in the state. That is certainly enough, but it could hardly be less efficient. It is really only for autonomous systems that the trivial realization is an efficient one. Note that there is also a *trivial past-induced* (take $X = \mathfrak{B}^-$) and a *trivial future-induced* realization (take $X = \mathfrak{B}^{+0}$).

2.2.2 *The past-induced canonical realization*

We will now introduce three equivalence relations on $\mathfrak{B}$ and show how these yield more efficient state representations. The first of these equivalence relations declares two elements of $\mathfrak{B}$ equivalent if, at time 0, they admit the same future continuations in $\mathfrak{B}$. It should be intuitively obvious that this captures very effectively the idea that two such trajectories bring the system in the same state at time 0: two copies of the same system are declared to be in the same state if we cannot think up any experiment which will give a different behaviour, a different future observation, in the first system versus the second. Thus we will call $w_1, w_2\in\mathfrak{B}$ *past equivalent*, denoted by $\simeq$, if $\{w_1 \underset{0^-}{\wedge} w\in\mathfrak{B}\} \Leftrightarrow \{w_2 \underset{0^-}{\wedge} w\in\mathfrak{B}\}$. It is clear that this defines an equivalence relation on $\mathfrak{B}$. Let $X^{\simeq} := \mathfrak{B}(\text{mod} \simeq)$ and define $\mathfrak{B}_s^{\simeq} \subseteq (W \times X^{\simeq})^T$ by $\{(w,x)\in\mathfrak{B}_s^{\simeq}\}:\Leftrightarrow\{w\in\mathfrak{B}$ and $x(t)=(\sigma^t w)(\text{mod} \simeq)$ for all $t\in T\}$.

PROPOSITION 2.1 $\Sigma_s^{\simeq} := (T, W, X^{\simeq}, \mathfrak{B}_s^{\simeq})$ defines a state representation of $\Sigma = (T, W, \mathfrak{B})$. It is called the *past-induced canonical state representation.*

Proof The external behaviour of $\Sigma_s^{\simeq}$ (with $X^{\simeq}$ viewed as a set of auxiliary variables) is clearly $\mathfrak{B}$. It remains to prove that $\Sigma_s^{\simeq}$ satisfies the axiom of state. Assume therefore that $(w_1, x_1), (w_2, x_2)\in\mathfrak{B}_s^{\simeq}$ have $x_1(0) = x_2(0)$, or, equivalently, that $w_1 \simeq w_2$. In order to prove the axiom of state we need to show that $(w_1, x_1)\Lambda_{0^-}(w_2, x_2)\in\mathfrak{B}_s^{\simeq}$, in other words, that

$$\sigma^t(w_1 \underset{0^-}{\Lambda} w_2) \simeq \begin{cases} \sigma^t w_1 & \text{for } t \le 0 \\ \sigma^t w_2 & \text{for } t > 0 \end{cases}.$$

Clearly $\{w_1 \simeq w_2\} \Rightarrow \{w_1 \Lambda_{0^-} w_2 \in \mathfrak{B}\}$, and the claim about the equivalence is trivial for $t \le 0$. We also have that $\{w_1 \simeq w_2\} \Rightarrow \{\sigma^t(w_1 \Lambda_{0^-} w) \simeq \sigma^t(w_2 \Lambda_{0^-} w)$ for all w and $t > 0\}$. Indeed, if $w_1 \simeq w_2$, then $\{w_1 \Lambda_{0^-} w \Lambda_{t^-} w' \in \mathfrak{B}\} \Leftrightarrow \{w_2 \Lambda_{0^-} w \Lambda_{t^-} w' \in \mathfrak{B}\}$. Applying this with $w = w_2$ yields the equivalence for $t > 0$. ■

2.2.3 *The future-induced canonical realization*

It may at first sight come as somewhat of a surprise that the backward version of the above construction yields a second equivalence relation which is in general different from the first one, but which will provide us with another state representation (forward in time).

We will call $w_1, w_2 \in \mathfrak{B}$ *future equivalent*, denoted by $\underset{\sim}{+}$, if $\{w\Lambda_{0^-} w_1 \in \mathfrak{B}\} \Leftrightarrow \{w\Lambda_{0^-} w_2 \in \mathfrak{B}\}$. It is clear that this again defines an equivalence relation on $\mathfrak{B}$. Let $X^{\underset{\sim}{+}} := \mathfrak{B}(\text{mod} \underset{\sim}{+})$ and define $\mathfrak{B}_s^{\underset{\sim}{+}} \subseteq (W \times X^{\underset{\sim}{+}})^T$ by $\{(w,x)\in\mathfrak{B}_s^{\underset{\sim}{+}}\} \Leftrightarrow \{w\in\mathfrak{B}$ and $x(t) = (\sigma^t w)(\text{mod} \underset{\sim}{+})$ for all $t\in T\}$. We have the following result:

PROPOSITION 2.2 $\Sigma_s^{\underset{\sim}{+}} := (T, W, X^{\underset{\sim}{+}}, \mathfrak{B}_s^{\underset{\sim}{+}})$ defines a state representation of $\Sigma = (T, W, \mathfrak{B})$. It is called the *future induced canonical state representation.*

Proof The proof of this proposition is fully analogous to that of Proposition 2.1 with the exception that the demonstration of the relevant equivalence is now trivial for $t \ge 0$ and requires proof for $t < 0$. Observe that $\{w_1 \underset{\sim}{+} w_2\} \Rightarrow \{\sigma^t(w\Lambda_{0^-} w_1) \underset{\sim}{+} \sigma^t(w\Lambda_{0^-} w_2)$ for all w and $t \le 0\}$. Indeed, if $w_1 \underset{\sim}{+} w_2$, then $\{w'\Lambda_{t^-} w \Lambda_{0^-} w_1 \in \mathfrak{B}\} \Leftrightarrow \{w'\Lambda_{t^-} w\Lambda_{0^-} w_2 \in \mathfrak{B}\}$. Now apply this with $w = w_1$. ■

The intuitive interpretation of the future induced canonical state representation is as follows. Two identical copies of a system which each produce a future trajectory are declared to be in the same state if they admit identical past histories compatible with these future observations.

2.2.4 *The two-sided canonical realization*

We will have occasion to use the following two-sided equivalence on $\mathfrak{B}$, which is the refinement of the partition of $\mathfrak{B}$ induced by the past and the future induced equivalences. We will call w_1, w_2 *two-sided equivalent*, denoted by $\overset{+}{\simeq}$, if $w_1 \simeq w_2$ and $w_1 \overset{+}{\sim} w_2$, in other words, if $\{w_1 \wedge_{0^-} w \in \mathfrak{B}\} \Leftrightarrow \{w_2 \wedge_{0^-} w \in \mathfrak{B}\}$ and $\{w \wedge_{0^-} w_1 \in \mathfrak{B} \Leftrightarrow \{w \wedge_{0^-} w_2 \in \mathfrak{B}\}$. It is clear that this once again defines an equivalence relation on $\mathfrak{B}$. Let $X^{\overset{+}{\simeq}} := \mathfrak{B}(\text{mod}\, \overset{+}{\simeq})$ and define $\mathfrak{B}_s^{\overset{+}{\simeq}} \subseteq (W \times X^{\overset{+}{\simeq}})^{\mathrm{T}}$ by $\{(w, x) \in \mathfrak{B}_s^{\overset{+}{\simeq}}\} :\Leftrightarrow \{w \in \mathfrak{B}$ and $x(t) = (\sigma^t w)(\text{mod}\, \overset{+}{\simeq})$ for all $t \in T\}$. We have the following result:

Proposition 2.3 $\Sigma_s^{\overset{+}{\simeq}} := (T, W, X^{\overset{+}{\simeq}}, \mathfrak{B}_s^{\overset{+}{\simeq}})$ defines a state representation of $\Sigma = (T, W, \mathfrak{B})$. It is called the *two-sided canonical state representation.*

Proof The proof of the proposition is a straightforward combination of the proof of Propositions 2.1 and 2.2. In order to show the axiom of state we need to show that $w_1 \overset{+}{\simeq} w_2$ implies (i) $w_1 \wedge_{0^-} w_2 \in \mathfrak{B}$; and

(ii)
$$\sigma^t(w_1 \underset{0^-}{\wedge} w_2) \overset{+}{\simeq} \begin{cases} \sigma^t w_1 & \text{for} \quad t \leq 0 \\ \sigma^t w_2 & \text{for} \quad t > 0 \end{cases}.$$

(i) is obvious. In order to prove (ii) and (iii), observe that $\{w_1 \overset{+}{\simeq} w_2\} \Rightarrow \{w_1 \simeq w_2\}$ which implies, by Proposition 2.1, $\sigma^t(w_1 \wedge_{0^-} w_2) \simeq \sigma^t w_1$ for $t \leq 0$ and $\sigma^t(w_1 \wedge_{0^-} w_2) \simeq \sigma^t w_2$ for $t > 0$. Repeating this for $\overset{+}{\sim}$ using Proposition 2.2, and combining both conclusions, yields the result. ∎

The above propositions yield three state constructions. The first one is based on the specification by which the past trajectory determines the future behaviour, the second is based on the specification by which a future trajectory allows to deduce the past behaviour, while the third is the combination of both.

As we shall see in the sequel important implications can be drawn from the fact the equivalence relations $\simeq$ and $\overset{+}{\sim}$ are equal. This is the case for autonomous systems and, more surprisingly, for linear systems.

Proposition 2.4 Let $\Sigma = (T, W, \mathfrak{B})$ be a dynamical system. Then, if Σ either is autonomous or linear, there holds: $\{w_1 \simeq w_2\} \Leftrightarrow \{w_1 \overset{+}{\simeq} w_2\} \Leftrightarrow \{w_1 \overset{+}{\sim} w_2\}$.

Proof (i) If Σ is autonomous, then $\{w_1 \simeq w_2\} \Leftrightarrow \{w_1^{0+} = w_2^{0+}\} \Leftrightarrow \{w_1 \overset{+}{\sim} w_2\}$ and the result is obvious.

(ii) If Σ is linear, observe first that $\{w_1 \simeq w_2\} \Leftrightarrow \{(w_1 - w_2)\Lambda_0\text{-}0 \in \mathfrak{B}\}$ and $\{w_1 \overset{+}{\sim} w_2\} \Leftrightarrow \{0\Lambda_0\text{-}(w_1 - w_2) \in \mathfrak{B}\}$. Hence $\{w_1 \simeq w_2\} \Leftrightarrow \{(w_1 - w_2)\Lambda_0\text{-}0 \in \mathfrak{B}\} \Leftrightarrow \{0\Lambda_0\text{-}(w_1 - w_2) \in \mathfrak{B}\} \Leftrightarrow \{w_1 \overset{+}{\sim} w_2\}$. ■

2.2.5 *An example*

We will now illustrate the above by means of a very simple example: a *pure delay*. Consider the system $\Sigma = (\mathbb{Z}, \mathbb{R}^2, \mathfrak{B})$ with $\mathfrak{B} = \{(w_1, w_2): \mathbb{Z} \to \mathbb{R}^2 \mid w_2(t) = w_1(t - \Delta), t \in \mathbb{Z}\}$. Here $\Delta \in \mathbb{Z}_+$ is a fixed number, the *length* of the delay.

Let us compute the equivalence relation $\simeq$ for this example. The equivalence of w' and w'' requires that the last Δ values of w_1' and w_1'' are equal. Hence $\{w' \simeq w''\} \Leftrightarrow \{w_1'(t) = w_1''(t)$ for $-\Delta \leq t < 0\}$. Define $x^{-}(t) = \text{col}\,[w_1(t-1), w_1(t-2), \ldots, w_1(t-\Delta)] =: \text{col}\,[x_1^{\simeq}(t), x_2^{\simeq}(t), \ldots, x_\Delta^{\simeq}(t)]$. The past induced canonical realization becomes $\Sigma_s^{-} = (\mathbb{Z}, \mathbb{R}^2, \mathbb{R}^\Delta, \mathfrak{B}_s^{\simeq})$ with $\mathfrak{B}_s^{\simeq} = \{(w, x^{\simeq}): \mathbb{Z} \to \mathbb{R}^2 \times \mathbb{R}^\Delta \mid w_1 = \sigma x_1^{\simeq}, w_2 = x_\Delta^{\simeq}$, and $E\sigma x^{\simeq} = F x^{\simeq}\}$, where

$$E := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \text{and } F := \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Now compute $\overset{+}{\sim}$. By considering the defining equation backwards in time, $w_1(t) = w_2(t + \Delta)$, we arrive at the equivalence $\{w' \overset{+}{\sim} w''\} \Leftrightarrow \{w_2'(t) = w_2''(t)$ for $0 \leq t \leq \Delta - 1\}$. Take hence $x^{\overset{+}{\sim}}(t) = \text{col}\,[w_2(t), w_2(t+1), \ldots, w_2(t + \Delta - 1)] =: \text{col}\,[x_1^{\overset{+}{\sim}}(t), x_2^{\overset{+}{\sim}}(t), \ldots, x_\Delta^{\overset{+}{\sim}}(t)]$. The future canonical realization becomes $\Sigma_s^{\overset{+}{\sim}} = (\mathbb{Z}, \mathbb{R}^2, \mathbb{R}^\Delta, \mathfrak{B}_s^{\overset{+}{\sim}})$ with $\mathfrak{B}_s^{\overset{+}{\sim}} = \{w, x^{\overset{+}{\sim}}): \mathbb{Z} \to \mathbb{R}^2 \times \mathbb{R}^\Delta \mid w_1 = \sigma x_\Delta^{\overset{+}{\sim}}, w_2 = x_1^{\overset{+}{\sim}}$, and $E x^{\overset{+}{\sim}} = F\sigma x^{\overset{+}{\sim}}\}$. The bilateral equivalence will lead to

$$\begin{aligned} x^{\overset{+}{\simeq}}(t) &= \text{col}\,[w_1(t-1), \ldots, w_1(t-\Delta), w_2(t), \ldots, w_2(t + \Delta^{-1})] \\ &=: \text{col}\,[x_1^{\overset{+}{\simeq}}(t), x_2^{\overset{+}{\simeq}}(t), \ldots, x_{2\Delta}^{\overset{+}{\simeq}}(t)] \end{aligned}$$

By constraining this vector by $x_{\Delta-i+1}^{\overset{+}{\simeq}} = x_{\Delta+i}^{\overset{+}{\simeq}}$ for $i = 1, 2, \ldots, \Delta$, we can consider $x^{\overset{+}{\simeq}}$ as an element of $\mathbb{R}^\Delta$. Now, the defining relation $w_2(t) = w_1(t - \Delta)$ shows that $x^{\simeq} \cong x^{\overset{+}{\sim}} \cong x^{\overset{+}{\simeq}}$ and the three equivalence relations, and hence the canonical realizations, will basically all be identical for this example (more precisely, the realizations are equivalent in the sense this will be defined in section 2.3.1). In fact, the state trajectories of these various realizations are related by $x^{\overset{+}{\sim}} = P x^{\overset{+}{\sim}}$ and $x^{\overset{+}{\simeq}} = \text{col}\,[x^{\overset{+}{\sim}}, P x^{\overset{+}{\sim}}] = \text{col}[P x^{\overset{+}{\sim}}, x^{\overset{+}{\sim}}]$, with

$$P := \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

2.3 More on the structure of state systems

2.3.1 *A partial order on and equivalence of realizations*

The above constructions have given us three trivial realizations and three canonical realizations. All these realizations are externally induced. From the second proposition in section 2.2.4 it follows also that the restrictions imposed on

$\mathfrak{B}$ by the requirement that the equivalence relations $\simeq$ and $\pm$ are equal is not as restrictive as could have been suspected at first sight. We shall see that it is precisely under this condition that efficient state representations are essentially unique. The key words here are *efficient* and *essentially*. We now set out to formalize this and introduce the required concepts.

We will address the question: *When should one system be considered equivalent to or more complex than another*? Our basic idea is to measure the complexity of a system by the 'number' of trajectories in its behaviour. We will see that this will allow us to formalize the fact that the canonical realizations cannot be simplified and, in particular, that they are simpler than the trivial realizations.

Let us consider two dynamical systems $\Sigma_1=(T,W,A_1,\mathfrak{B}_1)$ and $\Sigma_2=(T,W,A_2,\mathfrak{B}_2)$ with the same time set and external signal space but with possibly a different set of latent variables. We will call Σ_1 and Σ_2 *externally equivalent* if $P_w\mathfrak{B}_1=P_w\mathfrak{B}_2$, i.e., if they model the same external behaviour. However, if we want these systems to be essentially the same as models, it is reasonable to demand also that the latent variables should be related. This suggests calling Σ_1 and Σ_2 *equivalent* if there exists a bijection $f:A_1\to A_2$ such that $\{(w,a)\in\mathfrak{B}_1\}\Leftrightarrow\{(w,f\circ a)\in\mathfrak{B}_2\}$.

Now let $\Sigma=(T,W,\mathfrak{B})$ be a given dynamical system and denote by Σ_s all its state space representations. Thus, in the sense in which we have defined it above, all elements of Σ_s are externally equivalent. Specializing the notion of equivalence to Σ_s shows that two elements $\Sigma_i=(T,W,X_i,\mathfrak{B}_i)$, $i=1,2$, of Σ_s are equivalent if there exists a bijection $f:X_1\to X_2$ such that $\{(w,x_1)\in\mathfrak{B}_1\}\Leftrightarrow\{(w,f\circ z_1)\in\mathfrak{B}_2\}$. This obviously induces an equivalence relation on Σ_s. We will denote this equivalence by $\cong$. We will also introduce a pre-order on Σ_s. Let $\Sigma_s'=(T,W,X',\mathfrak{B}_s')$ and $\Sigma_s''=(T,W,X'',\mathfrak{B}_s'')$ be elements of Σ_s. Then $\{\Sigma_s''\leq\Sigma_s\}':\Leftrightarrow\{\exists$ surjection $f:X\to X'$ such that $\{(w,x'')\in\mathfrak{B}_s''\}\Rightarrow\{\exists(w,x')\in\mathfrak{B}_s'$ such that $x''=f\circ x'\}\}$ (in other words, if $\tilde{f}\circ\mathfrak{B}_s'\supseteq\mathfrak{B}_s''$, where $\tilde{f}:W\times X\to W\times X''$ is defined by $\tilde{f}(w,x'):=(w,f(x'))$.

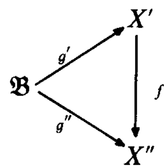
These definitions may be interpreted as follows. If $\Sigma_s'\geq\Sigma_s''$ then $\mathfrak{B}_s'$ contains at least as many trajectories as $\mathfrak{B}_s''$. It may contain more, for one thing, because $P_x\mathfrak{B}_s''$ may be a proper subset of $f\circ P_x\mathfrak{B}_s'$, or, when f is not injective, because certain trajectories in $\mathfrak{B}_s'$ may be represented in $\mathfrak{B}_s''$ more than once. If f is a bijection then it is logical to consider Σ_s' as equivalent to Σ_s'', since in that case their trajectories are in one-to-one correspondence: their state spaces are merely labelled differently.

This pre-order acquires more structure when we restrict attention to state trim externally induced representations. Firstly, it becomes a partial order modulo equivalence, and secondly, we can be a bit more specific about what the ordering means in this case.

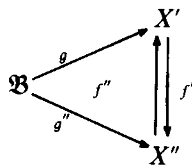
PROPOSITION 2.5 Let $\Sigma_s^{ext}:=\{\Sigma_s\in\Sigma_s\mid\Sigma_s$ is state trim and x is externally induced (see section 1.1.1. in Σ_s)$\}$. Then

(i) $\{\Sigma_s', \Sigma_s'' \in \Sigma_s^{\text{ext}}, \Sigma_s'' \leq \Sigma_s'\} \Leftrightarrow \{\exists$ a surjective map $f: X' \to X''$ such that $\{(w, x') \in \mathfrak{B}_s'\} \Rightarrow \{w, f \circ x' \in \mathfrak{B}_s''\}\}$, i.e., $\tilde{f} \circ \mathfrak{B}_s' = \mathfrak{B}_s''$;
(ii) $\{\Sigma_s', \Sigma_s'' \in \Sigma_s^{\text{ext}}, \Sigma_s' \leq \Sigma_s'' \leq \Sigma_s'\} \Rightarrow \{\Sigma_s' \cong \Sigma_s''\}$;
(iii) the trivial realization belongs to Σ_s^{ext} and is $\geq$ than every other element of Σ_s^{ext};
(iv) if $\Sigma_s \in \Sigma_s^{\text{ext}}$ and $\Sigma_s' \in \Sigma_s$ satisfy $\Sigma_s' \leq \Sigma_s$, then Σ_s' must belong to Σ_s^{ext}.

Proof First observe that if $\Sigma_s = (T, W, X, \mathfrak{B}_s) \in \Sigma_s^{\text{ext}}$, then there will exist a surjective map $g: \mathfrak{B} \to X$ such that $\{(w, x) \in \mathfrak{B}_s\} \Rightarrow \{x(0) = g(w)\}$, yielding $\{(w, x) \in \mathfrak{B}_s\} \Leftrightarrow \{x(t) = g(\sigma^t w)$ for all $t \in T\}$.
(i) Assume that $\Sigma_s' = (T, W, X', \mathfrak{B}_s')$, $\Sigma_s'' = (T, W, X'', \mathfrak{B}_s'')$ are elements of Σ_s^{ext} with $\Sigma_s'' \leq \Sigma_s'$. Let $f: X' \to X''$ be the map which expresses this ordering and $g': \mathfrak{B} \to X'$ and $g'': \mathfrak{B} \to X''$ the maps defined in the preamble. Then the following diagram



commutes. Let $(w, x') \in \mathfrak{B}_s'$. Now consider $(w, f \circ x')$. Then $(w, f \circ x') \in \mathfrak{B}_s''$, which yields (i).
(ii) By (i) there exist maps f', f'' such that the diagram



commutes. Hence $f \circ \tilde{f} = \mathrm{id}_{X''}$. It follows that $\Sigma_s' \cong \Sigma_s''$.
(iii) Take, in the commutative diagram in (i), $g' = \mathrm{id}_{\mathfrak{B}}$ and $f = g''$.
(iv) Let $\Sigma_s \in \Sigma_s^{\text{ext}}$, $\Sigma_s' \in \Sigma_s$, and $\Sigma_s' = (T, W, X', \mathfrak{B}_s') \leq \Sigma_s = (T, W, X, \mathfrak{B}_s)$. Then there exists a surjective map $f: X \to X'$ such that $(w, x') \in \mathfrak{B}_s'$ implies that there exists $(w, x) \in \mathfrak{B}_s$ such that $x' = f \circ x$. Let $g: \mathfrak{B} \to X$ be the surjective map which expresses the fact that Σ_s is externally induced. Now verify that $f \circ g$ is the sujective map which demonstrates that also Σ_s' is trim and externally induced. ■

The above proposition can be generalized in a straightforward manner to past and future induced realizations.

2.4 Are all minimal realizations equivalent?

2.4.1 *Minimal realizations and hidden variables*

Of special interest will be the *minimal state realizations*, defined as those elements $\Sigma_s^* \in \Sigma_s$ for which $\{\Sigma_s \in \Sigma_s, \Sigma_s \leq \Sigma_s^*\}$ implies $\{\Sigma_s \cong \Sigma_s^*\}$.

Whether a given realization $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ is or is not minimal may be difficult to check. As we shall see in the proposition which follows, a realization may not be minimal either because it is not trim or because it may contain states which can be lumped. Lumping is defined as follows. Let $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ be a state space system. Let X' be a subset of X, consider $X^{\text{red}} := (X')^{\text{complement}} \cup \{\alpha\}$, and define the map $l: X \to X^{\text{red}}$ specified by $l|_{(X')^{\text{complement}}} = \text{id}$ and $l(X') = \alpha$. We will say that the states in X' can be *lumped* if $(T, W, X^{\text{red}}, \mathfrak{B}^{\text{red}})$ still satisfies the axiom of state, where $\mathfrak{B}^{\text{red}} = \tilde{l}(\mathfrak{B}_s)$, and $\tilde{l}(w, x):(w, l\circ(x))$.

Examples of states which can be lumped are:

(i) Future-equivalent states in a past-induced realization; $x_1, x_2 \in X$ are said to be *future equivalent* if $(\mathfrak{B}(x_1))^+ = (\mathfrak{B}(x_2))^+$ where $\mathfrak{B}(x_0) := \{w \mid \exists x$ such that $(w, x) \in \mathfrak{B}_s$ and $x(0) = x_0\}$. Future equivalence defines an equivalence relation on X. Actually if Σ_s is a past induced state trim realization, then states can be lumped if and only if they are future equivalent.

(ii) *Past-equivalent* states (which are defined completely analogously) in a future induced realization.

Let $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ be a state space system. Define the *effective state space* as $X^{\text{eff}} := \{x \in X \mid \exists (w, x) \in \mathfrak{B}_s$ such that $x(0) = x\}$ and the *trimmed realization* as $\Sigma_s^{\text{trim}} := (T, W, X^{\text{eff}}, \mathfrak{B}_s)$. Obviously Σ_s is state trim iff $X = X^{\text{eff}}$ or, equivalently, iff $\Sigma_s = \Sigma_s^{\text{trim}}$. We have the following characterization of externally induced minimal realizations.

PROPOSITION 2.6 Assume that $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ is an externally induced state space representation of $\Sigma = (T, W, \mathfrak{B})$. Then it is a minimal representation if and only if Σ_s is state trim and the only non-empty sets of states which can be lumped are singletons.

Proof Consider $\Sigma_s^{\text{trim}} = (T, W, X^{\text{eff}}, \mathfrak{B}_s)$. Clearly $\Sigma_s^{\text{trim}} \leq \Sigma_s$. Observe, using Proposition 2.5 that $\Sigma_s \leq \Sigma_s^{\text{trim}}$ implies $X = X^{\text{eff}}$. Next, assume that $X' \subseteq X$ is a set of states which can be lumped. Consider $\Sigma_s^{\text{red}} = (T, W, X^{\text{red}}, \mathfrak{B}^{\text{red}})$ with X^{red} and $\mathfrak{B}^{\text{red}}$ defined as in the definition of lumping. Clearly $\Sigma_s^{\text{red}} \leq \Sigma_s$. Observe that $\Sigma_s \leq \Sigma_s^{\text{red}}$ implies, using Proposition 2.5, that X' is a singleton.

To prove the converse, assume that Σ_s is not minimal and that $\Sigma_s' = (T, W, X', \mathfrak{B}_s') \leq \Sigma_s$. By Proposition 2.5, Σ_s' must be trim and externally induced and there will exist a surjective map $f: X \to X'$ such that $\tilde{f} \circ \mathfrak{B}_s = \mathfrak{B}_s'$. Now observe that, since in X only singletons can be lumped, f must be injective. This allows the conclusion $\Sigma_s' \cong \Sigma_s$. ■

The state dynamics inside a set of states which can be lumped adds superfluous detail which is obviously not required in order to express and represent the external behaviour. Also, the existence of states outside X^{eff} can never be demonstrated by examining the external behaviour. Motivated by these considerations we will informally identify the non-minimality of a representation with the existence of what we will call *hidden* (state) *variables.*

2.4.2 *Two minimal state representations*

We have already met two minimal externally induced realizations:

THEOREM 2.1 *The canonical past-induced and the canonical future-induced realization are both minimal.*

Proof Since $\Sigma_s^{\simeq}$ is externally induced, it follows from Proposition 8 that it suffices to prove that $\Sigma_s^{\simeq}$ is state trim (which is obvious) and that the only non-empty sets which can be lumped are singletons. This however is a construction, $\Sigma_s^{\simeq}$ is past induced and contains no future equivalent states. The future-induced canonical realization is treated analogously. ■

2.4.3 *Equivalence of all minimal state representations*

We will now examine the question of *when all minimal state representations are equivalent.* As we shall see this is ***not always*** the case but it holds under the ***necessary and sufficient condition that the past-induced and the future-induced canonical realizations are identical.*** This very clean and general result implies for example that for autonomous systems and for linear systems all minimal state representations are indeed equivalent.

THEOREM 2.2 *Let* $\Sigma = (T, W, \mathfrak{B})$ *be a (time invariant) dynamical system and let* $\simeq$ *and* $\overset{+}{\sim}$ *denote the past, respectively future, induced equivalence on* $\mathfrak{B}$, *as introduced in sections 2.2.2 and 2.2.3. Then the following conditions are equivalent:*

(i) $\{w_1 \simeq w_2\} \Leftrightarrow \{w_1 \overset{+}{\sim} w_2\}$

(ii) $\{\Sigma_s', \Sigma_s'' \in \Sigma_s,\ \Sigma_s$ and Σ_s'' *both minimal*$\} \Rightarrow \{\Sigma_s' \cong \Sigma_s''\}$

In other words, all minimal realizations are equivalent if and only if $\mathfrak{B}(\text{mod} \simeq) = \mathfrak{B}(\text{mod} \overset{+}{\sim})$.

Proof (ii)⇒(i): By Theorem 2.1, $\Sigma_s^{\simeq}$ and $\Sigma_s^{\overset{+}{\sim}}$ are both minimal. If they are equivalent then there exists a bijection $f: X^{\simeq} \to X^{\overset{+}{\sim}}$ such that $w(\text{mod} \overset{+}{\sim}) = f \circ (w(\text{mod} \simeq))$. This implies that $\{w_1 \simeq w_2\} \Leftrightarrow \{w_1 \overset{+}{\sim} w_2\}$.

(i)⇒(ii): Let $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ be a state representation of $\Sigma = (T, W, \mathfrak{B})$. We will prove that $\{\Sigma_s$ minimal$\} \Rightarrow \{\Sigma_s \cong \Sigma_s^{\simeq} \cong \Sigma_s^{\overset{+}{\sim}}$N. Define $\mathfrak{B}(x) := \{w \in \mathfrak{B} \mid \exists \boldsymbol{x}$ such that $(w, \boldsymbol{x}) \in \mathfrak{B}_s$ and $\boldsymbol{x}(0) = x\}$. Then, by the state axiom, $\mathfrak{B}(x) =$

$(\mathfrak{B}(x))^- \wedge_0 (\mathfrak{B}(x))^{0+}$ and, of course, we also have $\mathfrak{B} = \bigcup_{x \in X} \mathfrak{B}(x)$. Now consider the analogous decomposition of $\mathfrak{B}$ as it is induced by the representation $\Sigma_s^{\simeq} = \Sigma_s^{\pm}$. Denoting its state space by A yields $\mathfrak{B} = \bigcup_{a \in A} \mathfrak{B}(a)$ with $\mathfrak{B}(a) = (\mathfrak{B}(a))^- \wedge_0 (\mathfrak{B}(a))^{0+}$. The fact that we are actually considering a representation which is at the same time the canonical past-induced and the canonical future-induced realization allows us to conclude that $\{a_1 \neq a_2\} \Rightarrow \{(\mathfrak{B}(a_1))^- \cap (\mathfrak{B}(a_2))^- = \emptyset$ and $(\mathfrak{B}(a_1))^{0+} \cap (\mathfrak{B}(a_2))^{0+} = \emptyset\}$. Furthermore, $\mathfrak{B} = \bigcup_{a \in A} \mathfrak{B}(a) = \bigcup_{x \in X} \mathfrak{B}(x)$. It is easy to see that this implies that for each $x \in X^{\text{eff}}$ there exists a (unique) $a \in A$ such that $\mathfrak{B}(x) = \mathfrak{B}^-(x)^- \wedge_0 (\mathfrak{B}(x))^{0+} \subseteq \mathfrak{B}^-(a) = \mathfrak{B}(a)$. This shows that there exists a decomposition of X^{eff}, the effective state space of X, into disjoint subsets X_a, $a \in A$, such that $\bigcup_{x \in X_a} B(x) = \mathfrak{B}(a)$. Now define $f: X \to A$ such that $f(X_a) = a$ and verify that $\{(w, x) \in \mathfrak{B}_s\} \Rightarrow \{(w, f \circ x) \in \mathfrak{B}_s^{\simeq}\}$. This shows that $\Sigma_s^{\simeq} \leq \Sigma_s$, and implies, using the definition of minimality, that all minimal state representations are equivalent to $\Sigma_s^{\simeq} = \Sigma_s^{\pm}$. ■

There are a large number of alternative equivalent ways of stating our necessary and sufficient condition for all minimal state representations of a given external behaviour to be equivalent. We now state some of them without further comments.

The following conditions are equivalent:

(1) *All minimal state representations are equivalent;*
(2) *Any one of the following three equalities is satisfied:*

$$\simeq \, = \, \underset{\sim}{+}; \quad \simeq \, = \, \underset{\simeq}{+}; \quad \underset{\sim}{+} \, = \, \underset{\simeq}{+};$$

(3) *The past-induced canonical realization is equivalent to the future-induced canonical realization;*
(4) *There exists a state representation which is at the same time past and future induced;*
(5) *If two pasts have one future in common, then they have all their futures in common. More precisely,* $\{w_1 \wedge_0 w \in \mathfrak{B}, w_2 \wedge_0 w \in \mathfrak{B}\} \Rightarrow \{w_1 \simeq w_2\}$;
(6) *If two futures have one past in common, then they have all their pasts in common. More precisely,* $\{w \wedge_0 w_1 \in \mathfrak{B},\ w \wedge_0 w_2 \in \mathfrak{B}\} \Rightarrow \{w_1 \underset{\sim}{+} w_2\}$.

Further, if any of the above conditions is satisfied then the unique minimal state representation $\Sigma_s^{\min} = (T, W, X^{\min}, \mathfrak{B}_s^{\min}) (\cong \Sigma_s^{\simeq} \cong \Sigma_s^{\pm})$ is both past and future induced and any other state representation $\Sigma_s = (T, W, X, \mathfrak{B})$ satisfies $\Sigma_s^{\min} \leq \Sigma_s$.

In fact, there will exist a surjective map $f: X^{\text{eff}} \to X^{\min}$ with X^{eff} the effective state space of Σ_s, such that $\tilde{f} \circ \mathfrak{B}_s = \mathfrak{B}_s^{\min} (\tilde{f}: W \times X^{\text{eff}} \to W \times X^{\min})$ is as usual define by $\tilde{f} \circ (w, x) = (w, f \circ x)$. The map f can be constructed as follows. Define, for $(w, x) \in \mathfrak{B}_s$, $f(x(0))$: $w(\text{mod} \simeq) = w(\text{mod} \underset{\sim}{+})$. Hence in this case the hidden variables in the non-minimal state representation Σ_s are precisely those in the complement of X^{eff} and in the equivalence classes X (mod ker f). The equivalence relation ker f is defined by $\{x_1 \ker f\, x_2\} :\Leftrightarrow \{f(x_1) = f(x_2)\}$.

As already mentioned, it is in general next to impossible to give conditions for a given state system to be a minimal state representation of its own external behaviour. However if this external behaviour satisfies any of the above conditions, then a state representation is minimal if and only if it is (i) trim; (ii) past induced and (iii) future induced. For linear systems, for instance, there are effective tests to verify these conditions.

All this implies that the belief, often *implicitly assumed*, that the state of dynamical system is a *uniquely* specified quantity, is not quite correct. For one thing, we should be considering minimal state representations and consider the state *up to equivalence* (*up to relabelling* of the state variables). More fundamentally, however, minimal state representations *may be intrinsically incomparable* (when $\simeq \neq \underset{\sim}{+}$). However, if $\simeq = \underset{\sim}{+}$ or, equivalently (see section 2.2.3), *if every past and every future experiment allows us to determine uniquely the present state, then the state is indeed essentially uniquely defined* (*modulo hidden variables and up to relabelling*). From the results in section 1.1.5 we can conclude that *all minimal state representations are equivalent for autonomous* (with $X \cong \mathfrak{B}$) *and for linear systems.*

2.4.4 *A smooth system with a non-unique minimal state*

In order to dispel the thought that this lack of non-uniqueness of minimal state representations is merely a matter of lack of smoothness, consider the system described by the following behavioural equations:

$$\dot{x} = f\circ(x, u); \quad y = r\circ(x, u); \quad w = (u, y).$$

Here $W = U \times Y, U = \mathbb{R}^m$, $Y = \mathbb{R}^p$, $X = \mathbb{R}^n$, $f:\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $r:\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$. These equations define a continuous time evolution law yielding the state space system $\Sigma_s = (\mathbb{R}, \mathbb{R}^m \times \mathbb{R}^p, \mathbb{R}^n, \mathfrak{B}_s)$ with $\mathfrak{B}_s := \{((u, y), x) | x$ is absolutely continuous $\dot{x}(t) = f(x(t), u(t))$ for almost all $t \in \mathbb{R}$, and $y(t) = r(x(t), u(t))$ for all $t \in \mathbb{R}\}$. Assume further that $\mathfrak{B}_s$ is a minimal state representation of its own external behaviour. Does it allow other non-equivalent minimal state representations? Or will this be essentially the unique minimal state representation if f and r are sufficiently smooth? In order to see that this is not necessarily the unique minimal state representation, assume that r takes the form $r(x, u) = \Sigma_{i=1}^m u_i r_i'(x, u)$ with $r_i':\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. Then *all* pasts have the future: $u(t) = 0, y(t) = 0$ for $t \geq 0$, in common. This shows that not all minimal state representations can be equivalent for this simple and yet quite general smooth nonlinear differential system.

2.4.5 *Past-induced minimal realizations*

We can of course also ask the question whether all past-induced minimal state representations are equivalent. Let $\Sigma = (T, W, \mathfrak{B})$ be a dynamical system and let Σ_s^{past} be its past-induced realizations, i.e., $\Sigma_s^{\text{past}} := \{\Sigma_s \in \Sigma_s | \Sigma_s$ is past induced$\}$. We state, without proof, the following result.

THEOREM 2.3

(i) *The canonical past-induced realization* $\Sigma_s^{\simeq} \in \Sigma_s^{\text{past}}$ *and is minimal;*
(ii) *All minimal realizations in* Σ_s^{past} *are equivalent;*
(iii) $\Sigma_s \in \Sigma_s^{\text{past}}$ *is minimal iff it is state trim and it contains no future equivalent states;*
(iv) $\{\Sigma_s \in \Sigma_s^{\text{past}}\} \Rightarrow \{\Sigma_s^{\simeq} \leq \Sigma_s \leq \Sigma_s^{\text{trivial/past}}\}$ *where* $\Sigma_s^{\text{trivial/past}}$ *denotes the trivial past-induced realization.*

Of course, a similar result is valid for future-induced realizations. The above theorem seems to make it appealing to concentrate on past-induced realizations. This assumption, moreover, sounds logical in that it may appear natural to consider the state as something which summarizes the *past* behaviour. However, it may be difficult, next to impossible, to verify that the state is indeed past induced even for very nice models (as for example the differential equation model of section 2.4.4). Also, the property of being past induced may be lost after interconnection of systems. As such, the assumption that the state is past induced is not really a natural one to impose as an axiom on state space systems. Note, however, that if the minimal state representation is essentially unique (see Theorem 2.2) then minimal implies past induced.

2.4.6 *Controllability*

The condition $\mathfrak{B}(\text{mod} \simeq) = \mathfrak{B}(\text{mod} \stackrel{+}{\sim})$ turns out to be a convenient assumption enabling us to prove a number of important theoretical results. We will pursue a few such items in the next three sections. We will first state a refinement of the second proposition in section 1.4.5, pertaining to the equivalence of controllability and point controllability of the induced state behaviour. We will delete the proof.

PROPOSITION 2.7 Let $\Sigma = (T, W, \mathfrak{B})$ satisfy $\mathfrak{B}(\text{mod} \simeq) = \mathfrak{B}(\text{mod} \stackrel{+}{\sim})$. Let $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ be a minimal state representation of Σ and $\Sigma_x = (T, X, \mathfrak{B}_x)$ be its state behaviour. Then Σ is controllable if and only if Σ_x is point controllable.

2.4.7 *Splitting variables*

Now consider the following question. Let $\Sigma = (T, W, \mathfrak{B})$ be the external behaviour of the system $\Sigma_a = (T, W, A, \mathfrak{B}_a)$ with latent variables. Assume that a splits w that is, that $\{(w_1, a_1) \in \mathfrak{B}_a, (w_2, a_2) \in \mathfrak{B}_a, a_1(0) = a_2(0)\} \Rightarrow \{(w_1, a_1) \Lambda_0 \text{-} (w_2, a_2) \in \mathfrak{B}_a\}$. Now view $\mathfrak{B} \subseteq W^{\mathrm{T}}$ as a subset of the product space $W^{(-\infty\, 0) \cap T} \times W^{[0, \infty) \cap T}$. Define $\mathfrak{B}(a) := \{w \in \mathfrak{B} \mid \exists a \text{ such that } (w, a) \in \mathfrak{B}_a \text{ and } a(0) = a\}$. The splitting property implies that $\mathfrak{B}(a) = (\mathfrak{B}(a))^- \Lambda_0 \text{-} (\mathfrak{B}(a))^{0+}$, hence $\mathfrak{B} = \bigcup_{a \in A} \mathfrak{B}(a) = \bigcup_{a \in A} (\mathfrak{B}(a))^- \Lambda_0 \text{-} (\mathfrak{B}(a))^{0+}$. We will call such a partition of $\mathfrak{B}$ a *minimal splitting* if the only non-empty sets $A' \subseteq A$ such that $\bigcup_{a \in A'} (\mathfrak{B}(a))^- \Lambda_0 \text{-} (\mathfrak{B}(a))^{0+} = \tilde{\mathfrak{B}}^- \Lambda_0 \text{-} \tilde{\mathfrak{B}}^{0+}$ for some $\tilde{\mathfrak{B}} - \subseteq \mathfrak{B}^-$ and $\tilde{\mathfrak{B}}^{0+} \subseteq \mathfrak{B}^{0+}$ are singletons. Now assume that the splitting $\mathfrak{B} = \bigcup_{a \in A} \mathfrak{B}(a)$ is minimal in this sense. *Will Σ_a also satisfy the axiom of state?* A partial answer is given in the following proposition, which we will not prove.

PROPOSITION 2.8 (i) Let $\Sigma_a = (T, W, A, \mathfrak{B})$ be a dynamical system with latent variables. Assume that the latent variables a are induced by either the strict past or the future of w. Then if the latent variables are minimal splitting, Σ_a is a state space system.
(ii) If $\Sigma = (T, W, \mathfrak{B})$, the external dynamical system induced by Σ_a, has the property that $\simeq\, = \overset{+}{\sim}$, and hence if all its minimal state representations are equivalent, then all its minimal splittings with latent variables define state space systems. Hence in this case minimal splitting and minimal state are equivalent.

2.4.8 *Evolution laws of minimal systems*

In Theorem 1.1 we have seen that a complete state space system $\mathfrak{B}_s$ can be faithfully represented by an evolution law. We have also seen that any system can be represented in state space form. The question thus arises. *Can a complete dynamical system be minimally realized by means of an evolution law?* The answer is in the affirmative.

THEOREM 2.4 *Let $\Sigma = (T, W, \mathfrak{B})$ be a time invariant complete dynamical system. Then*

(i) *if $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ is a state space representation of Σ, so is $\Sigma_s^{\text{complete}} = (T, W, X, \mathfrak{B}_s^{\text{complete}})$;*
(ii) *if, in addition, $\simeq\, = \overset{+}{\sim}$, then all minimal state space representations of Σ are complete.*

Proof (i) First observe that if $f: W_1 \to W_2$ is any map and if $\mathfrak{B}_1 \subseteq W_1^{\mathbb{T}}$ is complete, then $(f \circ \mathfrak{B}_1)^{\text{completion}} \supseteq f \circ \mathfrak{B}_1^{\text{completion}}$. Next observe that if $\mathfrak{B}_s$ satisfies the axiom of state, so does $\mathfrak{B}_s^{\text{completion}}$. Putting these two things together shows that $\Sigma_s^{\text{completion}}$ is a state space system with external behaviour equal to $\mathfrak{B}_s = \mathfrak{B}_s^{\text{completion}}$.
(ii) It suffices to prove that $\Sigma_s^{\simeq} \cong \Sigma_s^{\overset{+}{\sim}}$, the canonical past-induced realization of Σ, is complete. Consider $(\mathfrak{B}_s^{\simeq})^{\text{completion}}$. By (i), $(\mathfrak{B}_s^{\simeq})^{\text{completion}}$ defines also a state space representation of Σ. If $\mathfrak{B}_s^{\simeq}\, \mathrm{D} (\mathfrak{B}_s^{\simeq})^{\text{completion}}$ then there exist $w \in \mathfrak{B}$, $(w, x_1) \in \mathfrak{B}_s$, and $(w, x_2) \in (\mathfrak{B}_s^{\simeq})^{\text{completion}}$, with $x_1(0) \neq x_2(0)$. Now observe that $(\mathfrak{B}(w^-))^{0+} := \{w^{0+} \mid \bar{w} \wedge_0 w^{0+} \in \mathfrak{B})$ is equal to $(\beta(x_1(0)))^{0+} := \{w^{0+} \mid \exists (x, w) \in \mathfrak{B}_s$ such that $x(0) = x_1(0)\}$. Also, $(\tilde{\mathfrak{B}}(x_2(0)))^{0+} := \{w^{0+} \mid \exists (w, x) \in \mathfrak{B}_s^{\text{completion}}$ such that $x(0) = x_2(0)\}$, is included in $(\mathfrak{B}(w^{-1}))^{0+}$. Consequently $(\mathfrak{B}(x_2(0)))^{0+} := \{w^{0+} \mid \exists (x, w) \in \mathfrak{B}_s$ such that $x(0) = x_2(0)\} \subseteq (\mathfrak{B}(x_2(0)))^{0+} \subseteq (\mathfrak{B}(w^-))^{0+} = (\mathfrak{B}(x_1(0)))^{0+}$. Hence $(\mathfrak{B}(x_2(0)))^{0+} \subseteq (\mathfrak{B}(x_1(0)))^{0+}$. Since the minimal state representation is both past and future induced, this yields $x_1(0) = x_2(0)$. ■

It follows from Theorems 1.1 and 2.4(i) that a discrete time complete dynamical system can be faithfully represented by means of an evolution law. Using Theorem 2.4, the fact that the past-induced canonical realization is deterministic, and section 1.5.3, we conclude that a discrete time complete dynamical system admits a minimal state representation of the form

$$\sigma x = f \circ (x, w); \quad c \circ (x, w) = 0$$

If, moreover, $\simeq\, = \,\underset{\sim}{+}$, then all minimal state representations take this form, the only freedom remaining being a bijection on the state space X with the resulting modification of the maps f and c.

This shows that completeness, which from a systems point of view is a very reasonable hypothesis, is the crucial assumption which allows a dynamical system to be described by a set of first order difference equations.

2.4.9 *Discrete event systems*

Recall that we have called a time-invariant dynamical system $(\mathbb{Z}, W, \mathfrak{B})$ a *discrete event system* if W is a finite set, that is, if $|W| < \infty$ ($|\cdot|$ denotes the *cardinality*, that is $|W|$ equals the number of elements of W).

Let us consider a dynamical system $\Sigma = (\mathbb{Z}, W, \mathfrak{B})$ and assume that it has Δ-memory. Then a (past-induced) realization can be constructed as follows. Take $X = W^{\Delta}$ and $\mathfrak{B}_s$: $\{(w, x) \mid w \in \mathfrak{B}, x(t) = w|_{[t-\Delta, t)}\}$. This limits the cardinality of $X^{\simeq}$ (and similarly that of $X^{\underset{\sim}{+}}$) to the cardinality of W^{Δ}. Hence if a discrete even system is governed by a behavioural difference equation of lag L then it can be described by an evolution law with a *finite* state space, with $|X| \leq |W^{\Delta}|$. We will now examine the converse. We will meet the crucial condition of essential uniqueness of the state space also here!

THEOREM 2.5 *Let $\Sigma = (\mathbb{Z}, W, \mathfrak{B})$ be a discrete event system. Assume that $\mathfrak{B}$ is L-complete, equivalently that $\mathfrak{B}$ can be described by a difference equation of lag L. Then Σ can be realized by a discrete time evolution law $\Sigma_{\partial} = (T, W, X, \mathfrak{B}_s)$ with $|X| \leq |W^L|$. Conversely, if Σ can be realized by a discrete time evolution law with $|X| < \infty$ and if Σ has an essentially unique minimal state space realization (that is, if $\mathfrak{B}(\mathrm{mod} \simeq) = \mathfrak{B}(\mathrm{mod}\, \underset{\sim}{+})$), then $\mathfrak{B}$ can be described by a behavioural difference equation of lag $L \leq |X|(|X| - 1)/2$.*

Proof The first part of the theorem is clear from the preamble. To show the converse, observe first that the minimal state space representation of Σ' will have a state space containing at most $|X|$ elements. Let us therefore assume that Σ_{∂} is minimal. Since $\simeq\, = \,\underset{\sim}{+}$, ∂ will also be state deterministic. We will show that there exists a $\Delta \in \mathbb{Z}_+$ and a map $h: W^{\Delta} \to X$ such that $x(t) = h(w(t-1), \ldots, w(t-\Delta))$. Assume that (w', x'), $(w'', x'') \in \mathfrak{B}_{\partial}$ satisfy $w(t) = w''(t)$ for $0 \leq t < \Delta$. Consider the pairs $(x'(t), x''(t)) \in X^2$ for $0 \leq t \leq \Delta$. Assume that $x'(\Delta) \neq x''(\Delta)$. Observe that, by determinism, $\{x'(\Delta) \neq x''(\Delta)\} \Rightarrow \{x'(t) \neq x''(t)$ for $0 \leq t \leq \Delta\}$. It follows that if $\Delta \geq |X|(|X| - 1)/2$, there must exist $0 \leq t_1 < t_2 \leq \Delta$ such that $(x'(t_1), x''(t_1)) = (x'(t_2), x''(t_2))$. This implies that the periodic trajectory $w(t_1)w(t_1 + 1), \ldots, w(t_2 - 1)$ with period $(t_2 - t_1)$ is compatible with both the initial states x'_1 and x''_1. However, since $\simeq\, = \,\underset{\sim}{+}$, no two distinct initial states can have a common future w-trajectory emanating from it. Hence if $\Delta \geq |X|(|X| - 1)/2$, there exists a map $h: W^{\Delta} \to X$ such that $\{(w, x) \in \mathfrak{B}_{\partial}\} \Leftrightarrow \{w \in \mathfrak{B}$ and $\sigma^{\Delta} x = h \circ (\sigma^{\Delta - 1} w, \ldots, \sigma w, w)\}$. Now, let f be a map with domain W^{Δ} such that $f^{-1}(0) = \mathfrak{B}|_{[0, \Delta]} \subseteq W^{\Delta}$. Let $\mathfrak{B}^f$ be the behaviour induced by the difference equation $f(w(t), w(t+1), \ldots, w(t+\Delta)) = 0$. Clearly $\mathfrak{B} \subseteq \mathfrak{B}^f$. To show the converse, assume that $w \in \mathfrak{B}^f$ and define $x: \mathbb{Z} \to X$ by $x(t) = h(w(t - 1), \ldots, w(t - \Delta))$. Now verify that $(x(t), w(t), x(t+1)) \in \partial$ for all $t \in \mathbb{Z}$. Hence $\mathfrak{B} = \mathfrak{B}^f$ which proves the theorem. ■

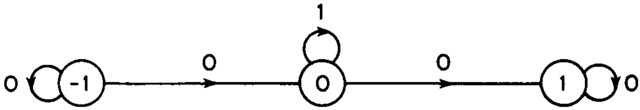


Fig. 10

Fig. 11

Note the possibilities which the difference equation constructed in the above theorem offers for error detection and correction. The discrete event system with three states shown in Fig. 10 demonstrates that the condition $\simeq\ =\ \underset{\sim}{+}$ is not superfluous for the possibility of describing a complete discrete event system by means of a difference equation.

The discrete event system of Fig. 10 does not have a unique minimal state representation. Note, however, that uniqueness of the minimal state representation is a sufficient but not a necessary condition enabling one to describe a behaviour by means of a difference equation. The nonequivalent discrete event system shown in Fig. 11 are both minimal state representations of the same external behaviour but can, however, be described by a simple difference equation.

2.4.10 *Recapitulation*

In section 1 we have seen that an evolution law induces a state space system and that a state space system induces an external behaviour. In this section we have studied the converse question. The results which we obtained may, in principle, allow us to write a given external behaviour as an evolution law, i.e., as a first order difference or differential equation. This involves first, finding a state space representation of a given behaviour and second, writing the state behaviour as an evolution law.

The problem of finding a state space representation of a given external behaviour is a very rich topic. There are two canonical representations which can always be constructed: the past-induced canonical realization and the future-induced canonical realization. Both, particularly the first, are based on an elegant and natural conceptualization of what really constitutes the state of a dynamical system. These realizations are minimal in the sense that they introduce no hidden variables, states whose dynamics are not reflected in the external behaviour. The question whether or not all minimal state representations are equivalent has, in general, a negative answer—even for very smooth systems. *As such the paradigm that barring irrelevancies (non-minimality–hidden variables) and up to relabelling (equivalence), the state is uniquely defined is, in general, not a valid one.* It is, however, valid if the past- and the future-induced canonical realizations coincide which is, in particular, true for autonomous and, more surprisingly, for linear systems.

In this section we have studied these questions in a set theoretic context. As such the development has acquired an unavoidable somewhat sterile flavour. Through the specialization to linear systems in section 4 we will make these representations more concrete and give this theory a more general appeal.

2.4.11 *Sources*

The state representation question as we defined it here was studied in [12] where the necessary and sufficient condition for uniqueness of the minimal state representation (Theorem 2.2) was first proven. Alternative versions and additional elements of this result were obtained in [13]. The problem of state representation and its importance to modelling were pioneered by Kalman [9], [16], particularly in the context of linear systems.

3 MODELS FROM OTHER MODELS—EXTRACTING THE CAUSE/EFFECT STRUCTURE OF A DYNAMICAL SYSTEM

In the approach which we have developed so far, all the components of the external attributes have played completely symmetric roles. This point of view is a logical one to start with: it allows us to accept any set of dynamical equations at face value and view them as the behavioural equations defining a dynamical system. However, in many situations some variables may cause other variables, some variables may be free, unconstrained, and should therefore be considered as unexplained by the model but imposed by the environment. We are, of course, thinking of the cause/effect, input/output (i/o) structure which may be present in a dynamical system. In view of the crucial role which i/o structures play in control and signal processing, such a refined classification of the external attributes is long overdue in our exposition. We will correct this situation in the present section. We will see in the next section that the input and output nature of certain attributes need not be imposed but can be deduced from the dynamical equations.

3.1 Inputs and outputs

In order to formalize input/output structures, we should give answers in our framework to the following questions:

(1) *When is one set of variables implied by another?*
(2) *When is one set of variables not anticipated by another?*
(3) *When is a set of variables free?*
(4) *What do we mean by inputs and outputs?*

In order to appreciate the definitions which follow, one should keep in mind that trajectories in the behaviour of a dynamical system will in general be generated on the one hand by an '*input*' signal—free signals unexplained by the model but imposed by the environment—and on the other hand by '*initial*' conditions—internal variables which have always been present, have been set up when the model was created at '$t = -\infty$', and which are hence also not explained by the model. Once the free input and the initial conditions are specified (and assuming, of course, that we know the dynamical laws of the system), we should be able to calculate the complete response of the dynamical system.

3.1.1 *Processing*

Let $\Sigma = (T, W_1 \times W_2, \mathfrak{B})$ be a dynamical system. Recall that $\mathfrak{B}_1 := P_{w_1}\mathfrak{B}$ and $\mathfrak{B}_2 := P_{w_2}\mathfrak{B}$. When $\mathfrak{B}$ is the graph of a map from $\mathfrak{B}_1$ to $\mathfrak{B}_2$, then we called w_2 observable from w_1. In that case w_1 completely specifies w_2. We are also interested in the case when w_1 specifies w_2 up to the initial conditions only. We will say that w_2 *processes* w_1 if $\{(w_1, w_2'), (w_1, w_2'')\in\mathfrak{B}; (w_2')^- = (w_2'')^-\} \Rightarrow \{w_2' = w_2''\}$. In other words, once w_1 is specified, the possible behaviour of w_2 is just like that of an autonomous system in which the past (or equivalently the internal initial conditions) completely specifies the future.

As an example, take the dynamical system described by a set of difference equations of the form

$$f_1(w_1(t+\Delta), w_1(t+\Delta-1), \ldots, w_1(t)) = 0$$
$$w_2(t) = f_2(w_2(t-1), \ldots, w_2(t-\Delta), w_1(t+\Delta), \ldots, w_1(t), \ldots, w_1(t-\Delta))$$

3.1.2 *Non-anticipation*

Consider $\Sigma = (T, W_1 \times W_2, \mathfrak{B})$. We will say that w_2 *does not anticipate* w_1 (or that w_1 *is not anticipated* by w_2) if $\{(w_1', w_2')\in\mathfrak{B}, w_1''\in\mathfrak{B}_1$, and $w_1'(t) = w_1''(t)$ for $t \le 0\} \Rightarrow \{\exists w_2''$ such that $(w_1'', w_2'')\in\mathfrak{B}$ and $w_2''(t) = w_2'(t)$ for $t \le 0\}$). In other words, if giving the strict future of w_1 in addition to its present and past does not provide additional information about the possible pasts and presents of w_2: it is the past, not the future of w_1 which influences the past of w_2. If $\{(w_1', w_2')\in\mathfrak{B}, w_1''\in\mathfrak{B}_1$, and $w_1'(t) = w_1''(t)$ for $t < 0\} \Rightarrow \{\exists w_2''$ such that $(w_1'', w_2'')\in\mathfrak{B}$ and $w_2''(t) = w_2'(t)$ for $t \le 0\}$, then we will say that w_1 is *strictly not anticipated by* w_2.

As an example, take the dynamical system described by a set of difference equations (or an analogous set of differential equations) of the form:

$$f_1(w_1(t+\Delta), w_1(t+\Delta-1), \ldots, w_1(t)) = 0$$
$$w_{21}(t) = f_2(w_{21}(t-1), \ldots, w_{21}(t-\Delta), w_{22}(t+\Delta'), \ldots, w_{22}(t), \ldots, w_{22}(t-\Delta), w_1(t), \ldots, w_1(t-\Delta))$$
$$w_2 = \begin{bmatrix} w_{21} \\ \cdots\cdots \\ w_{22} \end{bmatrix}$$

3.1.3 *Free variables*

Let $\Sigma = (T, W_1 \times W_2, \mathfrak{B})$ be a dynamical system. We will say that the variables w_1 are *free* if $\Sigma_1 = (T, W_1, P_{w_1}\mathfrak{B})$ is trim, memoryless, and complete. If $T = \mathbb{Z}$ then this means that $P_{w_1}\mathfrak{B} = (W_1)^{\mathbb{Z}}$, illustrating very well what the notion of '*being free*' expresses. We will call w_1 *locally free* if Σ_1 is trim and memoryless. Note that a locally free signal can be constrained at '$t = -\infty$' or '$t = +\infty$'. We will take locally free as an essential restriction on inputs.

3.1.4 *Input/output systems*

Our series of definitions culminates in the notion of input and output:

DEFINITION 3.1 An *input/output* (*i/o*) *dynamical system* is defined as a quadruple

$$\Sigma_{i/o} = (T, U, Y, \mathfrak{B})$$

with $T \subseteq \mathbb{R}$ the *time axis* (in this paper $T = \mathbb{R}$ or $\mathbb{Z}$); U the *input signal alphabet*; Y the *output signal alphabet*; and $\mathfrak{B} \subseteq (U \times Y)^{\mathrm{T}}$ the *behaviour*.

We postulate that $\mathfrak{B}$ (better: the induced dynamical system $(T, U \times Y, \mathfrak{B})$) satisfies the following axioms:

(A.1): u is locally free
(A.2): y processes u.

We will call $\Sigma_{i/o}$ a *non-anticipating i/o dynamical system* if, in addition:

(A.3): y does not anticipate u.

We will, as always in this paper assume also time invariance: $T = \mathbb{R}$ or $\mathbb{Z}$ and $\sigma^t\mathfrak{B} = \mathfrak{B}$ for all $t \in T$. We will call $\mathfrak{B}_u := P_u\mathfrak{B}$ (with $P_u : U \times Y \to U$ the projection $(u, y) \mapsto u$ and P_u considered as acting on U^{T} as well) the *input space*, elements of $\mathfrak{B}_u$ *inputs*, and those of $\mathfrak{B}_y := P_y\mathfrak{B}$ *outputs*.

The following theorem gives a reasonably concrete representation of non-anticipating i/o systems.

THEOREM 3.1 *Let* $\Sigma = (T, U \times Y, \mathfrak{B})$ *be a dynamical system. Then* $(T, U, Y, \mathfrak{B})$

defines a non-anticipating i/o system if and only if:

1. $\Sigma_u = (T, U, \mathfrak{B}_u)$ *is trim, and memoryless;*
2. $\forall (u^{-0}, y^{-0}) \in \mathfrak{B}^{-0}$, *there exists a non-anticipating map* $F: \mathfrak{B}_u^+ \to \mathfrak{B}_y^+$ *such that* $\{(u^{-0}, y^{-0}) \wedge_{0^+} (u^+, y^+) \in \mathfrak{B}\} \Leftrightarrow \{u^+ \in \mathfrak{B}_u^+$ *and* $y^+ = Fu^+\}$.

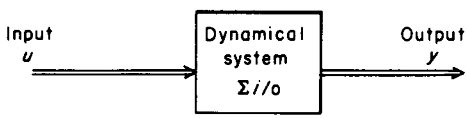
Proof (*if*): follows immediately from the definitions;

(*only if*): we will first show that for any given $(u^{-0}, y^{-0}) \in \mathfrak{B}^{-0}$ and $u^+ \in \mathfrak{B}_u^+$, there exists y^+ such that $(u^{-0}, y^{-0}) \wedge_{0^+} (u^+, y^+) \in \mathfrak{B}$. On the one hand, there exists $(\tilde{u}^+, \tilde{y}^+)$ such that $(u^{-0}, y^{-0}) \wedge_{0^+} (\tilde{u}^+, \tilde{y}^+) \in \mathfrak{B}$. On the other there exists $\hat{y}$ such that $(u^{-0} \wedge_{0^+} u^+, \hat{y}) \in \mathfrak{B}$. By non-anticipation this indeed implies that there exists y^+ such that $(u^{-0}, y^{-0}) \wedge_{0^+} (u^+, y^+) \in \mathfrak{B}$. Now consider, for a given $u^+ \in \mathfrak{B}_u^+$, those y^+ such that $(u^{-0}, y^{-0}) \wedge_{0^+} (u^+, y^+) \in \mathfrak{B}$. By the argument just given, there exists at least one such y^+. Since y processes u there is at most one such y^+. Hence there is a map $F: u^+ \mapsto y^+$. Now establish by a simple contradiction that this map must be non-anticipating. ■

Note that the above theorem gives us an alternative definition of a non-anticipating i/o system. However, we prefer to see locally free, processing, and non-anticipating as the defining properties and to view the above theorem as providing a convenient representation.

Input/output structures are usually displayed by means of *black-box signal flow graphs* as shown in Fig. 12.

Signal flow graphs constitute extremely useful, practical, transparent representations of dynamical systems. In particular they offer a compact notation for representing interconnections of subsystems (see Fig. 12). These black-box re-



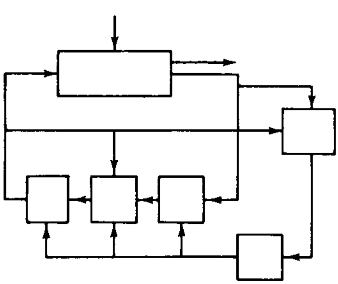


Fig. 12

presentations are far superior to what the standard mathematical operator notation has to offer, particularly when feedback (and hence implicit equations) is present in an interconnection. Black boxes visualize clearly the information processing features of systems and often preserve to some extent the physical lay-out as well.

3.1.5 *Causation*

When does one variable cause another? What do we mean by causation? From our definitions it follows that inputs are locally free, unconstrained, and such that they explain, together with the initial conditions and the dynamical laws, the signals produced by the system's behaviour. As such we define the input as a signal which can be viewed as imposed by the environment, as a signal which cannot be explained any further. If we have a non-anticipating input/output system then in addition the past of the output does not depend on the future of the input. In this sense we can think of a non-anticipating input/output system as formalizing a cause/effect relation. Thus, as we see it, in order to call something the cause we require that: (i) it cannot be explained or predicted by the phenomenon itself and (ii) it precedes, in time, the effect. Note that we consider the direction of time as an important element in our intuitive interpretation of cause/effect.

It is important not to read more into the notation of input than is claimed by its defining properties. An input can in principle be chosen freely by the modeller, and is compatible with the dynamical laws of the system. However, it does not mean that an input can always be viewed as a '*control*' variable. For example, the port voltages of an operational amplifier will be related as $w_2 = Aw_1$, with A the amplification factor. Now, if the voltage w_1 is applied at the input port, the voltage $w_2 = Aw_1$ will be realized at the output port. However, applying the voltage w_2 at the output port will not result in a voltage w_2/A appearing at the input port. Similarly, accepting a static relation between economic growth and inflation should not lead to the interpretation that imposing a certain inflation will cause a desired economic growth.

Of course, in conceptualizing vague notions as causation, there is an unavoidable degree of arbitrariness in the mathematical formalization. Some may want to view the fact that the output processes the input as the essential element, some may want to add the non-anticipation condition as being crucial, some may want to require that $\mathfrak{B}_u$ is not only trim and memoryless, but also complete. Note that trimness can always be achieved by redefining W_1 if need be.

Let us expand a little on the completeness condition in $\mathfrak{B}_u$. By suitably adapting Definition 3.1 to requiring 'free' instead of 'locally free' in (A.1), it is possible to prove a complete analogue of Theorem 3.1 with Σ_u trim, memoryless and complete. Requiring only locally free is useful when interpreting for example systems defined by convolution operators

$$w_2(t) = \int_{-\infty}^{+\infty} G(t-t')w_1(t')dt'$$

as (non-anticipating) i/o systems, where in order to interpret the right-hand side it may be necessary to require that w_1 has compact support or that it be square integrable (or square summable in the discrete time analogue).

3.1.6 *An example: Newton's second law*

In section 4 we will illustrate this series of definitions by means of systems described by high order linear difference or differential equations. In the present section we will see that even in the most common mechanical systems the choice of what is the input and what should be the output already presents a 'problem'.

(i) Consider Newton's second law

$$m\ddot{q} = F \qquad \text{(Nw1)}$$

relating the position q of a point mass with mass m to the force F exerted on it. Formally, $\Sigma = (\mathbb{R}, \mathbb{R}^3 \times \mathbb{R}^3, \mathfrak{B})$ with $\mathfrak{B} = \{(F, q): \mathbb{R} \to \mathbb{R}^3 \times \mathbb{R}^3 | F \in \mathscr{L}^{loc}(\mathbb{R}; \mathbb{R}^3),\ q \in \mathscr{C}^1(\mathbb{R}; \mathbb{R}^3),\ \dot{q}$ absolutely continuous, and $m\ddot{q}(t) = F(t)$ for almost all $t \in \mathbb{R}\}$. Let $\mathfrak{B}_F$ and $\mathfrak{B}_q$ be defined in the obvious way. Clearly $(\mathbb{R}, \mathbb{R}^3, \mathfrak{B}_F)$ is trim, memoryless, and complete. Further $(\mathfrak{B}((F,q)^{-0}))^+$ is given by the (F^+, q^+)'s satisfying $q^+(t) = q^{-0}(0) + \dot{q}^{-0}(0)t + (1/m)\int_0^t \int_0^\tau F^+(v)\, dv\, d\tau$. It follows from Theorem 3.1 that Σ defines a non-anticipating i/o system with F as input and q as output variable. It is easy to see on the other hand that $(\mathbb{R}, \mathbb{R}^3, \mathfrak{B}_q)$ is not memoryless and consequently that we could not have considered q as input instead.

$$F = -\frac{\partial V}{\partial \sigma}(q) \qquad \text{(Nw2)}$$

Formally $\Sigma = (\mathbb{R}, \mathbb{R}^3 \times \mathbb{R}^3, \mathfrak{B})$ with

$$\mathfrak{B} = \left\{(F, q): \mathbb{R} \to \mathbb{R}^3 \times \mathbb{R}^3 | F(t) = \frac{\partial V}{\partial \sigma}(q(t)) \quad \text{for all } t \in \mathbb{R}\right\}.$$

Obviously Σ defines a (static) dynamical system and it is easily seen that it is a non-anticipating input/output system with q the input variable and F the output variable. Since

$$\frac{\partial V}{\partial \sigma}: \mathbb{R}^3 \to \mathbb{R}^3$$

need not be a bijection it is in general not possible in this case to consider F to be the input variable instead.

(iii) A point mass which moves under the influence of the potential field V will be governed by the combination of (Nw1) and (Nw2), yielding

$$m\ddot{q} + \frac{\partial V}{\partial \sigma}(q) = 0$$

$$F = -\frac{\partial V}{\partial \sigma}(q) \qquad \text{(Nw3)}$$

This defines the dynamical system $\Sigma = (\mathbb{R}, \mathbb{R}^3 \times \mathbb{R}^3, \mathfrak{B})$ with $\mathfrak{B} = \{(F, q): \mathbb{R} \to \mathbb{R}^3 \times \mathbb{R}^3 | \text{(Nw3) is satisfied}\}$. Assume that V is sufficiently smooth so as to assure existence and uniqueness of solutions of the differential equation in (Nw3) for all initial conditions $(q(0), \dot{q}(0))$. The above system is an autonomous one: there are no inputs and both F and q should both be considered as outputs.

This example, however well known and on the beaten path, teaches us that what is the input and which is the output in a mechanical system will depend on the specific properties of the system and not just on the physical nature of the variables involved (even though we may intuitively—but unfortunately, incorrectly—tend to think of the force as being the input).

3.2 Input/state/output systems

3.2.1 *The structure of i/s/o systems*

It is possible to generalize all of the above concepts to state space systems $\Sigma_s = (T, U, X, \mathfrak{B}_s)$, in which we will view both x and y as being caused by u.

DEFINITION 3.2 An *input/state/output* (*i/s/o*) *dynamical system* is defined as a quintuple

$$\Sigma_{\mathrm{i/s/o}} = (T, U, Y, X, \mathfrak{B}_s)$$

with $T \subseteq \mathbb{R}$ the *time axis* (in this paper $T = \mathbb{R}$ or $\mathbb{Z}$); U the *input signal alphabet*; Y the *output signal alphabet*; X the *state space*; $\mathfrak{B}_s \subseteq (U \times Y \times X)^T$ the *behaviour* (we assume, of course, time-invariance: $\sigma^t \mathfrak{B}_s = \mathfrak{B}_s$ for all $t \in T$).

We postulate that $\mathfrak{B}_s$ (better: the induced dynamical system with latent variables $\Sigma_a = (T, U \times Y, X, \mathfrak{B}_s)$) satisfies the following axioms:

(A.1): $\mathfrak{B}_s$ satisfies the axiom of state;
(A.2): u is locally free;
(A.3): in $\mathfrak{B}_s$, (x, y) processes u;
(A.4): in $\mathfrak{B}_s$, u is strictly not anticipated by x and not anticipated by y.

From (A.1) it is clear that $(T, U \times Y, X, \mathfrak{B}_s)$ will be a state space system.

3.2.2 *Evolutive i/s/o structures.*

We will now introduce the classical dynamical systems in state form, governed by a state transition law and a read-out map.

DEFINITION 3.3 An *evolutive i/s/o dynamical structure* is defined as septuple

$$\Sigma_{\varphi,r} = (T, U, \mathcal{U}, Y, X, \varphi, r)$$

with $T \subseteq \mathbb{R}$ the *time axis* (in this paper $T = \mathbb{R}$ or $\mathbb{Z}$); U the *input alphabet*; $\mathcal{U} \subseteq U^T$ the input space; X the *state space*; Y the *output alphabet*; φ the *state transition law*; φ consists of a family of maps $\varphi_{t,u}$ from X into itself; one for each $t \in T$, $t > 0$, and for each $u \in \mathcal{U}|_{T \cap [0,t)}$, and $r: X \times U \to Y$ the *read-out map*.

We postulate that

(A.1): $\Sigma_u := (T, U, \mathscr{U})$ is locally free (and, of course, time invariant).

(A.2): φ satisfies the *semi-group property*:

$$\varphi_{t_2,u_2} \circ \varphi_{t_1,u_1} = \varphi_{t_1+t_2 u_1 * u_2}$$

with * *the concatenation-like product*, defined by

$$(u_1 * u_2)(t) := \begin{cases} u_1(t) & \text{for } 0 \leq t < t_1 \\ u_2(t) & \text{for } t \leq t_1 < t_1 + t_2 \end{cases}$$

The above, rather formidable, definition shows, through φ, how the state evolves under the influence of the input and, through r, how the output is generated from the state and the input. Thus $\varphi_{t,u}(x)$ is the *state reached under influence of the input u at time t starting from the initial state x*; while $r(x, u)$ denotes *the output value which will be generated when the system is in state x and an input with value u is applied to it*. Interpreting this leads to the *state space system induced by $\Sigma_{\varphi,r}$* defined as $\Sigma_s = (T, U \times Y, X, \mathfrak{B}_s)$, with

$$\mathfrak{B}_s = \{((u, y), x): T \to (U \times Y) \times X \mid u \in \mathscr{U};\ x(t_1) = \varphi_{t_1 - t_0, \sigma^{t_0} u|_{T \cap [0, t_1 - t_0)}}(x(t_0)),$$
$$\text{for all } (t_1, t_0) \in (T^2)_+;\ \text{and } y(t) = r(x(t), u(t)) \text{ for } t \in T\}.$$

Here $(T^2)_+ := \{(t_1, t_0) \in T^2 \mid t_1 \geq t_0\}$. It is easy to see that $\mathfrak{B}_s$ indeed satisfies the axiom of state. The external behaviour $\mathfrak{B}$ is then derived from $\mathfrak{B}_s$ in the usual way. Note that time-invariance has been built into the definition of φ and r. The behaviour $\mathfrak{B}_s$ can be viewed as being described by the behavioural equations

$$x(t + t') = \varphi_{t', \sigma^t u|_{T \cap [0, t')}}(x(t))$$
$$y(t) = r(x(t), y(t))$$

This is an infinite number of equations. However, each of the equations in the first category only involves x and u on the finite time interval $[t, t + t')$, while the equations in the second category are static equations. From this observation it follows immediately that $\mathfrak{B}_s$ is complete. We will see that the complete i/s/o systems are precisely those which can be described by means of systems $\Sigma_{\varphi,r}$ defined in terms of a state transition law and a read-out map. The surprising fact, perhaps, is the simple dependence of y on x and u which follows from the state and the other properties postulated of i/s/o systems.

THEOREM 3.2 (i) *Let $\Sigma_{i/s/o} = (T, U, Y, X, \mathfrak{B}_s)$ be a complete i/s/o dynamical system, and let $(T, U \times Y, X, \mathfrak{B}_s)$ be the state space system induced by it. Then there exists φ and r such that the evolutive i/s/o structure $\Sigma_{\varphi,r} = (T, U, P_u\mathfrak{B}_s, Y, X, \varphi, r)$ induces $(T, U, Y, X, \mathfrak{B}_s)$.*

(ii) *Conversely, let $\Sigma_{\varphi,r} = (T, U, \mathscr{U}, Y, X, \varphi, r)$ be an evolutive i/s/o structure, and* *let* $(T, U \times Y, X, \mathfrak{B}_s)$ *be the state space system induced by it. Then* $(T, U, Y, X, \mathfrak{B}_s)$ *defines a complete i/s/o dynamical system.*

Proof We will only give a broad outline of the proof.
(i) This part requires the construction of φ and r. Since x processes and strictly does not anticipate u, there exists, for all $(u, y, x) \in \mathfrak{B}_s$, a strictly non-anticipating map $F:(P_u\mathfrak{B}_s)^+ \to X^{T \cap (0,\infty)}$ which generates the future input/state trajectories. We will assume for simplicity that $\mathfrak{B}_s$ is state trim (the general case requires a separate construction of φ on the complement of X^{eff}). Now take any $x_0 \in X$, $t \in T, t > 0$, and $\tilde{u} \in P_u\mathfrak{B}_s|_{T \cap [0,t)}$ and define $\varphi_{t,\tilde{u}}(x_0) := x(t)$, with $x(t) = (Fu)(t)$, F the above constructed map corresponding to any $(u, y, x) \in \mathfrak{B}_s$ such that $x(0) = x_0$ and $u \in (P_u\mathfrak{B}_s)^+$ any input such that $u|_{T \cap [0,t)} = \tilde{u}$. Now verify that φ is well-defined and that it satisfies the semi-group property. In order to define $r(x, u)$ we need to prove that $\{(u_1, y_1, x_1), (\{(u_2, y_2, x_2) \in \mathfrak{B}_s, x_1(0) = x = x_2(0), u_1(0) = u = u_2(0)\} \Rightarrow \{y_1(0) = y_2(0) =: r(x, u)\}$. To see this, observe that $(u_1, y_1, x_1) \Lambda_{0^-} (u_2, y_2, x_2) \in \mathfrak{B}_s$ and use the fact that y processes and does not anticipate u. Next, verify that (φ, r) generates $\mathfrak{B}_s$.
(ii) The converse follows immediately from the definitions. ■

3.2.3 *Input/state/output evolution laws*

The most useful (and most common) dynamical models are those which actually combine all the advantages of the dynamical structures which we have considered up to now: they display their memory (the state), show their cause/effect (input/output) structure, and express their laws in a form which is purely 'local' in time (that is, as an evolution law). It is this class of models which is used most frequently in control applications.

DEFINITION 3.4 A *discrete time i/s/o evolution law* is defined as a sextuple

$$\Sigma_{\partial}^{i/s/o} = (T, U, X, Y, f, r)$$

with $T \subseteq \mathbb{R}$ the *time axis* (in this paper $T = \mathbb{Z}$); U, X, Y, and r as in Definition 3.3; and $f: X \times U \to X$ the *next state map*.

Its continuous time analogue reads

DEFINITION 3.5 A *continuous time i/s/o evolution law* is defined as a sextuple

$$\Sigma_{\partial}^{i/s/o} = (T, U, X, Y, f, r)$$

with $T \subseteq \mathbb{R}$ the *time axis* (in this paper $T = \mathbb{R}$); X a differentiable manifold, called the *state space*; U, Y and r as in Definition 3.3; and f the vector-field map; f is a map from U into the vector fields on X.

A discrete time i/s/o evolution law is thus described by the difference equation

$$\sigma x = f\circ(x, u); \quad y = r\circ(x, u),$$

while it continuous time counterpart is described by the differential equation

$$\dot{x} = f\circ(x, u); \quad y = r\circ(x, u)$$

The above definition is, clearly, a version of Definition 1.4 adapted to the i/o framework with $\{(x_0, (u, y), x_1) \in \partial\} \Leftrightarrow \{x_1 = f(x_0, u) \text{ and } y = r(x_0, u)\}$ in the discrete time case; and $\{((x, v), (u, y)) \in \partial\} \Leftrightarrow \{v = f(x, u) \text{ and } y = r(x, u)\}$ in the continuous time case. The state and external behaviour follow. The resulting input space $\mathscr{U}$ is defined as $\mathscr{U} = \{u: \mathbb{Z} \to U \mid \exists x: \mathbb{Z} \to X \text{ such that } \sigma x = f\circ(x, u)\}$ in the discrete time case and $\mathscr{U} = \{u: \mathbb{R} \to U \mid \exists x: \mathbb{R} \to X \text{ such that } \dot{x} = f\circ(x, u)\}$ in the continuous time case. Clearly $(T, U, \mathscr{U})$ is memoryless and shift invariant. We will assume that it is also trim (otherwise, simply redefine U) and complete. It is easy to see that this leads to an evolutive i/s/o system in the sense of Definition 3.3. The correspondence $f \to \varphi$ will undoubtedly be immediately clear from the nomenclature used. We will call the next state or vector field map f the *generator* of the corresponding state transition law φ.

3.2.4 *Maxwell's equations*

In section 1 we have already given a number of examples of i/s/o evolution laws. We will now illuminate *Maxwell's equations*, another *cause célébre* of the physical sciences, from this point of view.

Maxwell's equations relate the electric field, $\vec{E}(z, t)$, the magnetic field, $\vec{B}(z, t)$, the electric current density, $\vec{J}(z, t)$, and the electric charge density, $\rho(z, t)$; $z \in \mathbb{R}^3$ denotes the space coordinate and $t \in \mathbb{R}$ denotes time.

Maxwell's equations in free space read:

$$\nabla \cdot \vec{E} = \frac{\rho}{\varepsilon_0} \tag{EM1}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{EM2}$$

$$\nabla \cdot \vec{B} = 0 \tag{EM3}$$

$$c^2 \nabla \times \vec{B} = \frac{\vec{J}}{\varepsilon_0} + \frac{\partial \vec{E}}{\partial t} \tag{EM4}$$

Here $\nabla\cdot$ denotes the divergence, $\nabla\times$ the curl, ε_0 the dielectric constant of free space, and c the speed of light.

Maxwell's equations consist of two static (EM1 and EM3) and two dynamic (EM2 and EM4) behavioural equations. These laws (in particular EM1, EM3, and EM4) imply the law of conservation of charge:

$$\nabla \cdot \vec{J} = -\frac{\partial \rho}{\partial t} \tag{EM5}$$

How should we, as system theorists, look towards this system of equations? Obviously, they define a dynamical system with $T = \mathbb{R}$, $W = \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3) \times \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3) \times \mathscr{C}^0(\mathbb{R}^3; \mathbb{R}) \times \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3)$, and $\mathscr{B}$ described by the behavioural equations EM1 – EM4. In other words, we view Maxwell's equations as describing the behaviour of $(\vec{E}, \vec{B}, \rho, \vec{J})$.

These equations define an evolution law in the sense of Definition 1.4 with $T = \mathbb{R}$, $W = \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3) \times \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3) \times \mathscr{C}^0(\mathbb{R}^3; \mathbb{R}) \times \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3)$, $X = \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3) \times \mathscr{C}^1(\mathbb{R}^3; \mathbb{R})$ and

$$\partial = \{((\vec{x}_1, \vec{x}_2,), \vec{v}_1, \vec{v}_2,), \vec{w}_1, \vec{w}_2, w_3, \vec{w}_4 | \nabla \cdot \vec{x}_1 = \frac{w_3}{\varepsilon_0},$$

$$\nabla \cdot \vec{x}_2 = 0, \vec{v}_1 = c^2 \nabla \times \vec{x}_2 - \frac{\vec{w}_4}{\varepsilon_0}, \vec{v}_2 = -\nabla \times \vec{x}_1,$$

$$\vec{w}_1 = \vec{x}_1, \vec{w}_2 = \vec{x}_2$$

Note that we have taken the state space to be all of $\mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3) \times \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3)$ while (EM3) actually restricts the elements of the state space. This restriction has been incorporated in ∂ and leads to a system which, obviously, will not be state trim. Note that the choice of the state space is to a large extent free. In the above we have taken $(\vec{E}, \vec{B})$ as the state. We could have taken $(\vec{E}, \vec{B}, \rho, \vec{J})$ or $(\vec{E}, \vec{B}, \rho)$ (this last choice for example being suggested by considering EM1 – EM5).

If, however, we want to consider Maxwell's equations as an i/s/o evolution law in the sense of Definition 3.4, then we should incorporate (EM3) in the definition of the state space, yielding $T = \mathbb{R}$, $U = \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3)$, $y = \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3) \times \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3) \times \mathscr{C}^0(\mathbb{R}^3; \mathbb{R}) \times \mathscr{C}^0(\mathbb{R}^3; \mathbb{R}^3)$,

$$X = \{(x_1, x_2) \in \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3) \times \mathscr{C}^1(\mathbb{R}^3; \mathbb{R}^3) \nabla \cdot \vec{x}_2 = 0\},$$

and

$$f((\vec{x}_1, \vec{x}_2,), \vec{u}) \mapsto \left(c^2 \nabla \times \vec{x}_2 - \frac{\vec{u}}{\varepsilon_0}, -\nabla \times \vec{x}_1 \right), \quad \text{and} \quad r(x, \vec{u}) = (\vec{x}_1, \vec{x}_2, \nabla \cdot \vec{x}_1, \vec{u}).$$

Of course, by choosing $(\vec{E}, \vec{B}, \rho)$ constrained by EM1 and EM3, another trim i/s/o system will be obtained.

Thus the logical way of viewing (EM1 – EM4) is the following: $\vec{J}$ is the input, $(\vec{E}, \vec{B})$ subject to $\nabla \cdot \vec{B} = 0$ is the state, and ρ is an output determined by EM1. The state is trim and evolves according to EM2 and EM4. EM5 is an equation for the evolution of an output which can be deduced from the other equations. This state system is state trim and the state is (*trivially*) past and also future induced. Hence viewed this way we have recognized a minimal state realization of Maxwell's equations.

It is worthwhile to make a slightly philosophical comment. Maxwell's equations (EM1 – EM4) define an i/s/o system with $\vec{J}$ as input variable. It is reasonable and logical to consider each of these equations individually as a physical reality, regardless whether or not the others are (assumed to be) satisfied. In particular, one could assume EM1, EM2 and EM4 (with $\vec{J}$ and ρ as inputs) as a reality, without assuming EM3. In this case EM5 could not be derived. Feynman [17, pp. 18–3] basically argues that it makes no physical sense not to assume (EM5). From a system theoretic vantage point, we feel little sympathy for this point of view. *Is it really necessary to consider every time all the laws of physics all at once? Is it impossible to declare part of them as being true while disregarding the others?* This position is obviously untenable. The system theoretic, reductionist point of view, in which it is allowed to leave the environment $(\rho, \vec{B})$ totally unexplained (even though more scrutiny will undoubtedly lead to the discovery of additional relations: for example EM3), is much more reasonable. Any set of dynamical relations defines a dynamical system, a reality in its own right. It will usually contain unexplained inputs. Further analysis may lead to the discovery of more relations, leaving fewer unexplained inputs. This further analysis can, but need not be done. Otherwise we will always end up having to model the whole universe.

3.2.5 *State realization of i/o systems*

The construction of a state space representation for an i/o system can be approached using the methods developed in section 2 for general behaviours. In analogy with Theorem 3.1 it can be shown that $\Sigma_{i/o}=(T, U, Y, \mathfrak{B})$ defines an i/o system iff for each $(u^-, y^-)\in\mathfrak{B}^-$, there exists a map $F_{(u^-,y^-)}:\mathfrak{B}_u^{0+}\to\mathfrak{B}_y^{0+}$ which determines the future i/o pairs, i.e., this map is such that $\{u^-\Lambda_0\text{-}u^{0+}, y^-\Lambda_0\text{-}y^{0+}\in\mathfrak{B}\}\Leftrightarrow\{y^{0+}=F_{(u^-,y^-)}(u^{0+})\}$. Now consider the construction of the past induced canonical realization. This immediately shows that $\{(u_1,y_1)\simeq(u_2,y_2)\}\Leftrightarrow\{F_{(u_1^-,y_1^-)}=F_{(u_2^-,y_2^-)}\}$ Consequently the state in the canonical past induced realization stands in one-to-one relation to the future input/ouput map $F_{(u^-,y^-)}$.

An important special case of i/o dynamical systems are those in which the behaviour $\mathfrak{B}$ is given as the graph of a non-anticipating map $F:\mathfrak{B}_u\to Y^{\mathrm{T}}$ (called the *i/o map*). Consider the past-induced realization of this system. The past-induced equivalence realization $\simeq$ on $\mathfrak{B}$ now corresponds to what is called the *Nerode equivalence* $_{\tilde{N}}$ on $\mathfrak{B}_u^-$, defined as $\{u_1^-\,{}_{\tilde{N}}\,u_2\}:\Leftrightarrow\{F(u_1^-\Lambda_0\text{-}u^{0+})=(F(u_2^-\Lambda_0\text{-}u^{0+}))^{0+}$ for all $u^{0+}\in\mathfrak{B}^{0+u}\}$. The state x is thus seen to be generated by a map $\mathfrak{B}_u\to X^{\mathrm{T}}$ which is strictly non-anticipating in the sense that $x(0)$ is generated by a map $F_x:\mathfrak{B}_u^-\to X$ such that $\{((u,y),x)\in\mathfrak{B}_s^-\}\Rightarrow\{x(0)=F_x(u^-)\}$. Note that *state trim* now means what is classically (at least when we assume that all trajectories start from a common *ground state*) referred to as *state reachability* (meaning that F_x is surjective).

Specializing Theorem 2.3 to i/o systems show that the realization which takes $F_{(u^-,y^-)}$ as its state (or the Nerode equivalence class in the case of i/o maps) will be minimal and that all minimal past-induced realizations will be equivalent to it. Finally, all minimal realizations will be equivalent for an i/o system iff this past-induced realization is also future induced. This property requires that x be observable from the future of (u, y), meaning that any future i/o pair $(u^{0+}, y^{0+})\in\mathfrak{B}^{0+}$ will determine the present state uniquely.

3.2.6 *The construction of the input space*

In section 2 we have solved the problem of associating with any dynamical system a state space realization. The key element in the solution of this problem was the construction of a suitable state space. We have seen, at least in principle, how this construction can be carried out. The analogous problem in the context of the concepts developed in the present section would be the construction of an i/o representation of a given behaviour. We will give a very rough idea of how this question can be approached. This problem will be pursued in more concrete terms in the next section for linear systems. The key element now will be the construction of the input alphabet. For simplicity of presentation we will discuss this problem only for complete discrete time systems. We have seen in section 1.5.3 that such systems can always be represented by means of a deterministic evolution law

$$\sigma x = \tilde{f} \circ (x, w); \quad c \circ (x, w) = 0.$$

The problem is to write this evolution law as in i/s/o evolution law.

If we are lucky (as will be the case for the linear systems of section 4) there may exist a decomposition $W = U \times Y$ and a map $r: X \times U \to Y$ such that $\{c(x,(u,y)) = 0\} \Leftrightarrow \{y = r(x,u)\}$. This yields the i/s/o representation

$$\sigma x = f \circ (x, u); \quad y = r \circ (x, u).$$

with $f(x,u) := \tilde{f}(x,(r(x,u))$, u the input variable, y the output variable, and $w = (u, y)$. A more natural approach, however, is to look for a set U, a map $j: W \to U$, and a map $r: X \times U \to W$ such that $\{w \mid c(x,w) = 0\} = \{\text{im}\, r(x,.)\}$ for all $x \in X$. Further, set $Y = W$, and consider the representation

$$\sigma x = f \circ (x, u), \quad w = r \circ (x, u);$$

with $f(x,u) : \tilde{f}(x, r(x,u))$ as an i/s/o evolution law. Hence $u = j(w)$ can then be considered the input variable and $y = w$ as the output variable. Here u should be regarded as a latent variable introduced in order to explain the w-trajectories in the behaviour as being generated by a free input signal. It is easy to verify that, on the set theoretic level discussed here, such an input alphabet U always exists. This approach recognizes $j(w)$ as the free input part of w. In general, however, we prefer to view the construction of a universal input alphabet U as a problem which is a little artificial. Alternatively, we may want to let the map j be also dependent on x, yielding

$$\sigma x = f \circ (x, u), \quad w = h \circ (x, u), \quad \text{and} \quad u = j \circ (x, w).$$

It is, indeed, much more reasonable to look upon the free input variable as a state dependent object, defined as the elements of the set $\{w \mid c(x,w) = 0)$. Thus at each instant of time we consider the 'input' as the part of the w-variable which can be chosen freely: as such it will be determined by the present state. The joint input/state space has then the structure of a bundle with base space X and with the fibre above x playing the role of the free input alphabet when the system is in state x.

This elementary discussion is obviously also valid for continuous time evolution laws described by $\dot{x} = \tilde{f} \circ (x, w)$; $c \circ (x, w) = 0$.

3.2.7 *Addition*

Many of the concepts and model classes introduced in the first three sections of this paper can be illustrated by means of an example from elementary arithmetic: the addition of two real numbers. Consider the ordinary decimal expansion of $\boldsymbol{a} \cong \sum_{t=-\infty}^{+\infty} \boldsymbol{a}(t)10^t$ and identify the real numbers $\boldsymbol{a}$ with the time series $\boldsymbol{a}: \mathbb{Z} \to D := \{0, 1, \ldots, 9\}$. The real numbers $\boldsymbol{b}$ and $\boldsymbol{c}$ will be treated similarly. Then the relation induced by addition defines a dynamical system: $\Sigma = (\mathbb{Z}, D^3, \mathfrak{B})$ with $\mathfrak{B} = \{(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}): \mathbb{Z} \to D^3 \mid \exists t_1$ such that $\boldsymbol{a}(t) = \boldsymbol{b}(t) = \boldsymbol{c}(t) = 0$ for $t > t_1$ and $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}\}$. It is easy to see that this dynamical system is time invariant.

Consider the state space realization problem. The partition of $\mathfrak{B}$ generated by the past-induced canonical realization is easy to identify. There are two equivalence classes, denoted by $\{0, 1\}$. There holds:

$$\{(a, b, c) \in 0\} \Leftrightarrow \left\{ \sum_{t=-\infty}^{-1} a(t)10^t + \sum_{t=-\infty}^{-1} b(t)10^t < 1 \right\}$$

and

$$\{(a, b, c) \in 1\} \Leftrightarrow \left\{ \sum_{t=-\infty}^{-1} a(t)10^t + \sum_{t=-\infty}^{-1} b(t)10^t \geq 1 \right\}.$$

This leads to a state space realization where we identify the state trajectory with the binary sequence $x: \mathbb{Z} \to \{0, 1\} =: B$, yielding the realization $\Sigma_s = (\mathbb{Z}, D^3, B, \mathfrak{B}_s)$ with $\mathfrak{B}_s = \{((a, b, c), x): \mathbb{Z} \to D^3 \times B \mid (a, b, c) \in \mathfrak{B}$ and

$$x(t) = \begin{Bmatrix} 0 \text{ if } \sum_{t=-\infty}^{-1} a(t)10^t + \sum_{t=-\infty}^{-1} b(t)10^t < 1 \\ 1 \text{ if } \sum_{t=-\infty}^{-1} a(t)10^t + \sum_{t=-\infty}^{-1} b(t)10^t \geq 1 \end{Bmatrix}.$$

Now compute the evolution law induced by this system. This yields $\partial = \{(x_0, (a, b, c), x_1) \mid x_1 = (x_0 + a + b)\, \mathrm{DIV}\,(10)$ and $c = (x_0 + a + b)\, \mathrm{MOD}\,(10)\}$.† Hence the trajectories in $\mathfrak{B}_s$ can be recursively computed as:

$$x(t+1) = (x(t) + a(t) + b(t))\, \mathrm{DIV}\,(10)$$
$$c(t) = (x(t) + a(t) + b(t))\, \mathrm{MOD}\,(10).$$

These equations should be familiar ones: we have rederived the handy algorithm for addition which was taught to us in the first grade of elementary school. This algorithm was put into its present form in the sixteenth century by *Simon Stevin* who was also born in Bruges in Flanders and spent most of his professional life in the service of the Dutch government. It is enlightening to observe that *Stevin's algorithm is* basically nothing more than *a direct implementation of realization theory and finding the associated evolution law.* It follows from Theorem 1.1 that this evolution law will actually realize the completion of $\mathfrak{B}$: because of the condition $a(t) = b(t) = c(t) = 0$ for t sufficiently large, $\mathfrak{B}$ itself is not complete.

Let us now examine the i/o structure of $\mathfrak{B}$. It is easily seen that $\oplus: (a, b) \mapsto c = a + b$ defines a non-anticipating i/o map from $\mathcal{U} = \{(a, b): \mathbb{Z}^2 \to D^2 \mid \exists t_1$ such that $a(t) = b(t) = 0$ for $t \geq t_1\}$ into $D^{\mathbb{Z}}$. This yields the non-anticipating i/o system $\Sigma_{i/o} = (\mathbb{Z}, D^2, D, \mathfrak{B}_{i/o})$ with $\mathfrak{B}_{i/o} = \{((a, b), c) \mid a, b \in \mathcal{U};\ c = a + b\}$. Similarly, it is possible to the behaviour $\mathfrak{B}_s$ defined above as an (evolution) i/s/o system.

Now, turn to the question of the uniqueness of the minimal realization. It is easy to see that for the case at hand, $\simeq\ =\ \underset{\sim}{+}$, and hence the state constructed above corresponds to the unique minimal state realization. Hence it is also the canonical future-induced realization. As such it should be possible to run Stevin's algorithm backwards. This yields

$$x(t-1) = 10\, x(t) - a(t-1) - b(t-1) + c(t-1)$$

subject to the constraint

$$x(t) = (10\, x(t) + 2a(t) + 2b(t) - c(t))\, \mathrm{DIV}\,(10)$$

† Let $n, m, d, r \in \mathbb{Z}_+$, $m > 0$, be such that $n = dm + r$, $0 \leq r < m$. Then $d =: (n)\, \mathrm{DIV}\,(m)$ and $r =: (n)\, \mathrm{MOD}\,(m)$.

However, this backwards recursion can only be brought in i/s/o form by regarding the driving input variable as a latent variable: no combination of the variables (a, b, c) will serve as input variables for this backward recursion. It is for good reasons that Stevin's algorithm runs so successfully the way it does.

3.7.8 *Recapitulation*

In this section we have shown how the input/output–cause/effect structure can be incorporated into the framework set forward in Definition 1.1. We have taken as essential features of input, first, that it is locally free: that is, that it has no local structure of its own, and second, that the output processes the input, that is, that the input, together with the initial conditions, completely determines the response of the system. In many applications, it is important to incorporate non-anticipation. Combining non-anticipating input/output structures with the notion of state yields the input/state/output systems of Definition 3.2. In an i/s/o system the state satisfies the state axiom and processes the locally free input while it does not strictly anticipate the input. We have seen that i/s/o systems can always be described by a state transition law and a read-out map. This leads to the notion of i/s/o evolution law in which the state evolution is governed by a next state map. This structure, which displays both the cause/effect and the memory structure, and which expresses the behavioural equations in a one-step recursive form, is a particularly useful one in applications.

3.7.9 *Sources*

Many of the definitions in the beginning of this section are given here for the first time. Initial versions appear in [14]. Definition 3.3 and realization theory for input/output systems is extensively discussed in [9].

4 LINEAR TIME INVARIANT SYSTEMS—MODELS OF ALL SHAPES AND IN ALL SIZES

In this section we will treat a very important class of dynamical systems: the linear time invariant complete systems. This family of systems has been studied a great deal both in control and in electrical circuit theory, and it forms the theoretical basis for bread and butter applications in control engineering, signal processing, and econometrics. As we shall see, these systems admit a surprisingly simple mathematical characterization: in discrete time, their behaviour corresponds precisely to the closed linear shift invariant subspaces of $(\mathbb{R}^q)^{\mathbb{Z}}$, equipped with the topology of pointwise convergence. We will prove that this class of systems are also those which can be described by a set of behavioural equations consisting of a finite number of recursive linear equations (we shall call these *(AR)-equations*) or equivalently the time invariant systems which admit a linear evolution law with an underlying state space which is finite-dimensional. We will also see that the external attributes of such systems can always be partitioned into two sets: one set consists of input variables—they act as arbitrary causes imposed by the environment on the system; the other set consists of output variables—they act as effects, produced, through the system dynamics, by the inputs and the internal initial conditions.

4.1 Polynomial operators in the shift

4.1.1 *Polynomials and polynomial matrices*

As usual $\mathbb{R}[s]$ denotes the polynomials with real coefficients in the indeterminate s. We will also consider 'polynomials' with both positive and negative powers in s. These are sometimes called *dipolynomials*—we will simply call them polynomials. We will denote those as $\mathbb{R}(s,s^{-1}]$; $\mathbb{R}[s^{-1}]$ consists of the elements containing non-positive powers only. The vector and matrix analogues are written as $\mathbb{R}^n[s]$, $\mathbb{R}^{n_1\times n_2}[s]$, $\mathbb{R}^n[s,s^{-1}]$, $\mathbb{R}^{n_1\times n_2}[s,s^{-1}]$, etc. Of course, an element of $\mathbb{R}^{n_1\times n_2}[s,s^{-1}]$ can be considered either as a matrix of polynomials or as a polynomial with matrix coefficients. We will not make any such distinction: sometimes the first interpretation is the more natural one, sometimes the second is.

Both $\mathbb{R}[s]$ and $\mathbb{R}[s,s^{-1}]$ are rings, with the obvious definition of addition and multiplication. The *unimodular* (i.e. invertible) elements of $\mathbb{R}[s]$ are the non-zero constants, while in $\mathbb{R}[s,s^{-1}]$ they are the non-zero monomials, i.e., the elements of the form αs^d, $\alpha \neq 0$. Thus elements of $\mathbb{R}^{n\times n}[s]$ are *unimodular* (i.e. invertible as polynomial matrices) iff their determinant equals a non-zero constant, while elements of $\mathbb{R}^{n\times n}[s,s^{-1}]$ are invertible iff their determinant equals αs^d for some $d\in\mathbb{Z}$ and $\alpha\neq 0$. $\mathbb{R}(s)$ denotes the real rational functions in the indeterminate s; $\mathbb{R}(s)$ is the fraction field of $\mathbb{R}[s]$, of $\mathbb{R}[s,s^{-1}]$, and of $\mathbb{R}[s^{-1}]$. If $f(s)\in\mathbb{R}(s), f(s)=q(s)/p(s)$ with $p(s), q(s)\in\mathbb{R}[s]$ and degree $p\geq$ degree q, then we will call f *proper*. The collection of all proper rational functions will be denoted by $\mathbb{R}_+(s)$. If degree $p>$ degree q then we call the rational function *strictly proper*. Further, $\mathbb{R}_-(s):=\{f(s)\in\mathbb{R}(s)\,|\,f(s^{-1})\in\mathbb{R}_+(s)\}$. The spaces $\mathbb{R}^n(s)$, $\mathbb{R}^n_+(s)$, $\mathbb{R}^n_-(s)$, etc. are analogously defined. When referring to ranks, determinants, or minors of elements of $\mathbb{R}^{n_1\times n_2}(s)$, $R^{n_1\times n_2}[s]$, or $\mathbb{R}^{n_1\times n_2}[s,s^{-1}]$, we will consider them as matrices over the field $\mathbb{R}(s)$. However, when considering elements of $\mathbb{R}(s)$, $\mathbb{R}[s]$ or $\mathbb{R}[s,s^{-1}]$, it is sometimes useful to identify the indeterminate s with a complex number $\lambda\in\mathbb{C}$. As such speaking about the rank of an element of $M(s)\in\mathbb{R}^{n_1\times n_2}(s)$, $\mathbb{R}^{n_1\times n_2}[s]$, or $\mathbb{R}^{n_1\times n_2}[s,s^{-1}]$ may become a little ambiguous. In order to avoid this, we have sometimes used the notation $\mathit{rank}_{\mathbb{R}(s)}M(s)$ when we consider the rank of M viewed as a matrix of rational functions, and $\mathit{rank}_{\mathbb{C}}M(\lambda)$ when we consider λ as an element of $\mathbb{C}$ and evaluate the rank of the complex matrix $M(\lambda)$. A similar notation applies to polynomials and polynomial matrices $M(s,s^{-1})$, and for im or ker.

An important result which we will frequently use is the *Smith form* of a polynomial matrix. Consider a matrix polynomial $R(s)\in\mathbb{R}^{n_1\times n_2}(s)$. Then by pre- and post-multiplication by unimodular matrices $U_1(s)\in\mathbb{R}^{n_1\times n_1}[s]$ and $U_2(s)\in\mathbb{R}^{n_2\times n_2}[s]$, $R(s)$ can be brought in Smith form:

$$U_1(s)R(s)U_2(s)=\left[\begin{array}{c:c}\mathrm{diag}\,[d_1(s),d_2(s),\ldots,d_r(s)] & 0\\ \hdashline 0 & 0\end{array}\right]$$

with $d_i(s) \in \mathbb{R}[s]$, $d_i \neq 0$, $i = 1, 2, \ldots, r$, and d_{i+1} a factor of d_i for $i = 1, 2, \ldots, r-1$. A similar statement holds, *mutatis mutandis*, for $R[s, s^{-1}] \in \mathbb{R}^{n_1 \times n_2}[s, s^{-1}]$.

4.1.2 *Sequence spaces*

Let $\mathbb{L}^q$ denote the space of time series $w: \mathbb{Z} \to \mathbb{R}^q$, equipped with the topology of pointwise convergence: $\{w_n \xrightarrow[n \to \infty]{} w\} :\Leftrightarrow \{w_n(t) \xrightarrow[n \to \infty]{} w(t)$ for all $t \in \mathbb{Z}\}$ (this last convergence should be understood in the norm topology). $\mathbb{L}^q$ is a separable, metrizable topological space. We will denote by $\mathscr{L}^q$ the collection of all linear, closed, shift invariant subspaces of $\mathbb{L}^q$. This space (as behaviours of systems) will play an extremely important role in this section. The *backwards shift* $\sigma: \mathbb{L}^q \to \mathbb{L}^q$ is, as usual, defined by $(\sigma f)(t) := f(t+1)$. Obviously σ defines an invertible continuous linear operator on $\mathbb{L}^q$, with σ^{-1} the *forward shift*, $(\sigma^{-1} f)(t) = f(t-1)$. Now consider the polynomial matrix $R(s, s^{-1}) \in \mathbb{R}^{n_1 \times n_2}[s, s^{-1}]$, $R(s, s^{-1}) = R_L s^L + R_{L-1} s^{L-1} + \cdots + R_{l+1} s^{l+1} + R_l s^l$. Then the operator $R(\sigma, \sigma^{-1})$ from $(\mathbb{R}^{n_2})^{\mathbb{Z}}$ to $(\mathbb{R}^{n_1})^{\mathbb{Z}}$ given by $(R(\sigma, \sigma^{-1}) f)(t) := R_L f(t+L) + R_{L-1} f(t+L-1) + \cdots + R_{l+1} f(t+l+1) + R_l f(t+l)$ defines a continuous linear operator from $\mathbb{L}^{n_2}$ to $\mathbb{L}^{n_1}$. We will call such an operator a *polynomial operator in the shift.*

4.1.3 *The action of polynomial operators in the shift on $\mathbb{L}^n$*

In this section we will answer the following questions. *What subsets can be written as kernels of polynomial operators in the shift? As images? When is a polynomial operator in the shift injective? Surjective? Bijective? How do polynomial operators in the shift transform closed linear shift invariant subspaces?* We will state our results in a sequence of propositions. First, however, we will prove a lemma which is of some interest in its own right.

LEMMA Let $R(\sigma, \sigma^{-1}): \mathbb{L}^{n_2} \to \mathbb{L}^{n_1}$ be a polynomial operator in the shift. Then $R(\sigma, \sigma^{-1})$ is:

(i) $\{\text{injective}\} \Leftrightarrow \{\text{rank}_{\mathbb{C}} R(\lambda, \lambda^{-1}) = n_1$ for all $0 \neq \lambda \in \mathbb{C}\}$;
(ii) $\{\text{surjective}\} \Leftrightarrow \{\text{rank}_{\mathbb{R}(s)} R(s, s^{-1}) = n_2\}$;
(iii) $\{\text{bijective}\} \Leftrightarrow \{n_1 = n_2$ and $R(s, s^{-1})$ is unimodular$\}$.

Proof If $U(s, s^{-1}) \in \mathbb{R}^{n \times n}[s, s^{-1}]$ is unimodular with $N(s, s^{-1}) = (U(s, s^{-1}))^{-1}$, then $U(\sigma, \sigma^{-1}) N(\sigma, \sigma^{-1}) = N(\sigma, \sigma^{-1}) U(\sigma, \sigma^{-1}) = \text{id}_{\mathbb{L}^n}$, showing that $U(\sigma, \sigma^{-1})$ is indeed a bijection. Now consider any $R(s, s^{-1}) \in \mathbb{R}^{n_1 \times n_2}[s, s^{-1}]$. Let $URV = \Lambda$ be the Smith form of R. Bijectivity of $U(\sigma, \sigma^{-1})$ and $V(\sigma, \sigma^{-1})$ implies that it suffices to prove (i) and (ii) for R's which are of the form

$$\left[\begin{array}{c:c} \text{diag}\,[d_1(s), d_2(s), \ldots, d_r(s)] & 0 \\ \hdashline 0 & 0 \end{array}\right].$$

Observe that it hence suffices to prove that $d(\sigma, \sigma^{-1}): \mathbb{L}^1 \to \mathbb{L}^1$ is (i) injective iff d is unimodular, and (ii) surjective iff $d \neq 0$. If $d(s, s^{-1}) = \alpha s^d$, $\alpha \neq 0$, then $d(\sigma, \sigma)^{-1}$ is bijective, hence injective. If $d(s, s^{-1}) = r_L s^L + \cdots + r_l s^l$ with $L > l$, $r_L r_l \neq 0$, then there obviously exists a non-zero solution of the difference equation

$$r_L w(t+L)+\cdots+r_l w(t+l)=0 \qquad t\in\mathbb{Z}$$

(simply choose $w(L-1),\ldots,w(l)$ arbitrary and extend these initial conditions forward and backwards by solving the difference equation). Hence if $d(s,s^{-1})$ is not unimodular, $d(\sigma,\sigma^{-1})$ is not injective.

Now consider the surjectivity of $d(\sigma,\sigma^{-1})$. Clearly $d\neq 0$ is a necessary condition. To show that it is also sufficient, let $d(s,s^{-1})=r_L s^L+\cdots+r_l s^l$ have $r_L r_l\neq 0$. We will show that for any $f:\mathbb{Z}\to\mathbb{R}$, there exists a solution to the difference equation

$$r_L w(t+L)+\cdots+r_l w(t+l)=f(t) \qquad t\in\mathbb{Z}$$

In order to see this, choose $w(L-1),\ldots,w(l)$ arbitrary and define recursively

$$w(t)=\begin{cases}\dfrac{1}{r_L}(f(t-L)-r_{L-1}w(t-1)-\cdots-r_l w(t-(L-l))) & \text{for } t\geq L\\ \dfrac{1}{r_l}(f(t-l)-r_L w(t+(L-l)))-\cdots-r_{l+1}w(t-l+1)) & \text{for } t<l\end{cases}$$

■

The following proposition is the first of many representation results which will be obtained in this section. It shows that $\mathscr{L}^q$ consists precisely of the kernels of the polynomial operators in the shift.

PROPOSITION 4.1A $\{\mathfrak{B}\in\mathscr{L}^q\}\Leftrightarrow\{\exists g\in\mathbb{Z}_+$ and $R(s,s^{-1})\in\mathbb{R}^{g\times q}[s,s^{-1}]$ such that $\mathfrak{B}=\ker R(\sigma,\sigma^{-1})\}$. Moreover, we can always take $0\leq g\leq q$, and take R to be a polynomial matrix $R(s)\in\mathbb{R}^{g\times q}[s]$ or, for that matter, $R(s^{-1})\in\mathbb{R}^{g\times q}[s^{-1}]$. If $g=0$, define $\ker R(\sigma,\sigma^{-1})=\mathbb{L}^q$.

Proof ($\Leftarrow$): $\ker R(\sigma,\sigma^{-1})$ is obviously linear. It is also shift invariant. Indeed, $\{R(\sigma,\sigma^{-1})w=0\}\Leftrightarrow\{R(\sigma,\sigma^{-1})\sigma w=0\}$. To show that $\ker R(\sigma,\sigma^{-1})$ is closed, observe that $\{w_n\xrightarrow[n\to\infty]{}w,(R(\sigma,\sigma^{-1})w_n)(t)=0\}\Rightarrow\{R(\sigma,\sigma^{-1})w(t)=0\}$. ($\Rightarrow$): $\mathbb{L}^q$ and $\mathbb{R}^{1\times q}[s,s^{-1}]$ are duals with, for $a\in\mathbb{R}^{1\times q}[s,s^{-1}]$ and $b\in\mathbb{L}^q$, $\langle a,b\rangle:=(a(\sigma,\sigma^{-1})b)(0)$. Duality should be interpreted in the sense that $\mathbb{R}^{1\times q}[s,s^{-1}]$ specifies the continuous linear functionals on $\mathbb{L}^q$. Now $\mathbb{R}^{1\times q}[s,s^{-1}]$ is a finitely generated free module over the principal ideal domain $\mathbb{R}[s,s^{-1}]$. This implies that every submodule is finitely generated. Now consider, for $\mathfrak{B}\in\mathscr{L}^q$, its dual, $\mathfrak{B}^{\mathrm{T}}:=\{a\in\mathbb{R}^{1\times q}[s,s^{-1}]\,|\,\langle a,b\rangle=0\}$. By shift invariance of $\mathfrak{B}$, $s\mathfrak{B}^\perp=\mathfrak{B}^\perp$. Hence $\mathfrak{B}^\perp$ is a submodule or $\mathbb{R}^{1\times q}[s,s^{-1}]$, implying that there exist $r_1(s,s^{-1}),\ldots,r_g(s,s^{-1})\in\mathbb{R}^{1\times q}[s,s^{-1}]$ such that $\mathfrak{B}^\perp=\mathbb{R}[s,s^{-1}]r_1(s,s^{-1})+\cdots+\mathbb{R}[s,s^{-1}]r_g(s,s^{-1})$. Define $(\mathfrak{B}^\perp)^\perp:=\{w:\mathbb{Z}\to\mathbb{R}^q\,|\,\langle r,w\rangle=0$ for all $r\in\mathfrak{B}^\perp\}$. Since $\mathbb{L}^q$ is a topologial vector space $(\mathfrak{B}^\perp)^\perp$ is the smallest closed linear subspace containing $\mathfrak{B}$. Hence, since $\mathfrak{B}$ is closed, $\{w \in \mathfrak{B}\} \Leftrightarrow \{w \perp \mathfrak{B}^\perp\}$. Consequently, using shift-invariance, $\mathfrak{B} = \{w: \mathbb{Z} \to \mathbb{R}^q \mid r_1(\sigma, \sigma^{-1})w = 0, r_2(\sigma, \sigma^{-1})w = 0, \ldots, r_g(\sigma, \sigma^{-1})w = 0\}$. Now define $R(s, s^{-1}) = \mathrm{col}\,[r_1(s, s^{-1}), r_2(s, s^{-1}), \ldots, r_g(s, s^{-1})]$ and conclude that $\mathfrak{B} = \ker R(\sigma, \sigma^{-1})$. This proof also demonstrates that we can take $0 \leq g \leq q$ (since $\mathfrak{B}^\perp$ is a submodule of $\mathbb{R}^{1 \times q}[s, s^{-1}]$, which is a q-dimensional module over $\mathbb{R}[s, s^{-1}]$). That R can be taken to be a polynomial follows from that fact that, for all $d \in \mathbb{Z}$, $\ker R(\sigma, \sigma^{-1}) = \sigma^d \ker R(\sigma, \sigma^{-1}) = \ker \sigma^d R(\sigma, \sigma^{-1})$. ■

A more elementary proof of the implication ($\Rightarrow$) of the above proposition can be found in [14a].

Our second result identifies the subsets of $\mathbb{L}^q$ which are images of polynomial operators in the shift. Let $\mathfrak{B} \subseteq (\mathbb{R}^q)^{\mathbb{Z}}$. We will denote by $\mathfrak{B}^{\text{compact}}$, $\mathfrak{B}^{\text{closure}}$, and $\mathfrak{B}^{\text{compact/closure}}$ respectively the set of elements of $\mathfrak{B}$ with compact support, the closure of $\mathfrak{B}$, and the closure of $\mathfrak{B}^{\text{compact}}$, i.e., $\mathfrak{B}^{\text{compact/closure}} := \{w: \mathbb{Z} \to \mathbb{R}^q \mid \exists\, w_n \in \mathfrak{B}$ such that w_n has compact support and $w_n \xrightarrow[n \to \infty]{} w\}$.

Proposition 4.1B $\{\mathfrak{B} \in \mathcal{L}^q$ and $\mathfrak{B} = \mathfrak{B}^{\text{compact/closure}}\} \Leftrightarrow \{\exists\, m \in \mathbb{Z}_+$ and $M(s, s^{-1}) \in \mathbb{R}^{q \times m}[s, s^{-1}]$ such that $\mathfrak{B} = \mathrm{im}\, M(\sigma, \sigma^{-1})\}$. Moreover, we can always take $0 \leq m \leq q$, and take M to be a polynomial matrix $M(s) \in \mathbb{R}^{q \times m}[s]$ or, for that matter, $M(s^{-1}) \in \mathbb{R}^{q \times m}[s^{-1}]$. If $m = 0$, define $\mathrm{im}\, M(\sigma, \sigma^{-1}) = 0$.

Proof ($\Rightarrow$): Since $\mathfrak{B} \in \mathcal{L}^q$, $\mathfrak{B} = \ker R(\sigma, \rho^{-1})$ for a suitable $R(s, s^{-1}) \in \mathbb{R}^{g \times q}[s, s^{-1}]$. Let us now examine what the condition $\mathfrak{B} = \mathfrak{B}^{\text{compact/closure}}$ signifies about R. We claim that it is equivalent to $\mathrm{rank}_{\mathbb{C}}\, R(\lambda, \lambda^{-1}) = \mathrm{rank}_{\mathbb{R}(s)} R(s, s^{-1})$ for all $0 \neq \lambda \in \mathbb{C}$, in other words, that the Smith form of R is equal to

$$R = U \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} V.$$

In order to see this, observe that with T unimodular $T(\sigma, \sigma^{-1})\mathfrak{B}^{\text{compact/closure}} = (T(\sigma, \sigma^{-1})\mathfrak{B})^{\text{compact/closure}}$. Now let $R = U \Lambda V$ with

$$\Lambda = \begin{bmatrix} \mathrm{diag}\,[d_1, \ldots, dr] & 0 \\ 0 & 0 \end{bmatrix}$$

be the Smith form of R. Then $\mathfrak{B} = \ker R(\sigma, \sigma^{-1})$ is given by $\mathfrak{B} = V^{-1}(\sigma, \sigma^{-1}) \ker \Lambda(\sigma, \sigma^{-1})$. Hence, in the obvious notation, with $e_i := \mathrm{col}\,[0, \ldots, 0, 1, 0, \ldots 0]$ and the 1 in the ith position, $\ker \Lambda(\sigma, \sigma^{-1}) = e_1 \ker d_1(\sigma, \sigma^{-1}) \oplus \cdots \oplus e_r \ker d_r(\sigma, \sigma^{-1}) \oplus e_{r+1} \mathbb{L}^1 \oplus \cdots \oplus e_q \mathbb{L}^1$. Now (examine the proof of the injective part of the above lemma) if $d \neq 0$, $\ker d(\sigma, \sigma^{-1}) = 0$ iff d is unimodular and otherwise $\ker d(\sigma, \sigma^{-1})$ is finite dimensional with no non-zero elements having compact support. Hence if all the d_i's are unimodular, $(\ker \Lambda(\sigma, \sigma^{-1}))^{\text{compact/closure}} = e_{r+1} \mathbb{L}^1 \oplus \cdots \oplus e_q \mathbb{L}^1$. We conclude that $\{(\ker R(\sigma, \sigma^{-1}))^{\text{compact/closure}} = e_{r+1} \mathbb{L}^1 \oplus \cdots$

$\oplus e_q\mathbb{L}^1\} \Leftrightarrow \{\operatorname{rank} R_{\mathbb{C}}(\lambda, \lambda^{-1}) = \operatorname{rank}_{\mathbb{R}(s)} R(s, s^{-1})$ for all $0 \neq \lambda \in \mathbb{C}\} \Leftrightarrow$

$$\left\{ R(s, s^{-1}) = U(s, s)^{-1} \begin{bmatrix} I_r & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix} V(s, s^{-1}) \text{ with } U, V \text{ unimodular} \right\}.$$

Hence

$$\{\mathfrak{B} = \mathfrak{B}^{\text{compact/closure}}\} \Leftrightarrow \{\mathfrak{B} = \ker [0 \vdots I_{q-r}] V(\sigma, \sigma^{-1})\}$$

$$\Leftrightarrow \left\{ \mathfrak{B} = \operatorname{im} V^{-1}(\sigma, \sigma^{-1}) \begin{bmatrix} I_q \\ 0 \end{bmatrix} \right\}.$$

($\Leftarrow$): Consider the Smith form of M, $M = U\Lambda V$. Hence $\operatorname{im} M(\sigma, \sigma^{-1}) = U(\sigma, \sigma^{-1}) \operatorname{im} \Lambda(\sigma, \sigma^{-1})$. Applying the lemma yields

$$\operatorname{im} M(\sigma, \sigma^{-1}) = \operatorname{im} U(\sigma, \sigma^{-1}) \begin{bmatrix} I_r \\ 0 \end{bmatrix} = \ker [0 \vdots I_{q-r}] U^{-1}(\sigma, \sigma^{-1}).$$

This implies $\operatorname{im} M(\sigma, \sigma^{-1}) \in \mathscr{L}^q$ and (see the proof of (i)) $\operatorname{im} M(\sigma, \sigma^{-1}) = (\operatorname{im} M(\sigma, \sigma^{-1}))^{\text{compact/closure}}$. That we can take $0 \leq m \leq q$ and $M(s, s^{-1}) = M_1(s) \in \mathbb{R}^{q \times m}[s]$ or $M(s, s^{-1}) = M_2(s^{-1}) \in \mathbb{R}^{q \times m}[s^{-1}]$ is obvious. ■

The proposition which follows recognizes in particular a general family of linear operators on an infinite-dimensional topological space which map certain closed subspaces into closed subspaces.

Proposition 4.1C Let $R(s, s^{-1}) \in \mathbb{R}^{q_1 \times q_2}[s, s^{-1}]$. Then the polynomial operator in the shift $R(\sigma, \sigma^{-1})$ maps elements of $\mathscr{L}^{q_2}$ into elements of $\mathscr{L}^{q_1}$. Conversely, the inverse image under $R(\sigma, \sigma^{-1})$ of an element of $\mathscr{L}^{q_1}$ is an element of $\mathscr{L}^{q_2}$.

Proof We will prove the converse part first. Let $\mathfrak{B}_1 \in \mathscr{L}^{q_1}$. By Proposition 4.1A there exists an $R_1(\sigma, \sigma^{-1})$ such that $\mathfrak{B}_1 = \ker R_1(\sigma, \sigma^{-1})$. Now $(R(\sigma, \sigma^{-1}))^{-1}\mathfrak{B}_1 = \ker R_1(\sigma, \sigma^{-1}) R(\sigma, \sigma^{-1})$, which, by Proposition 4.1A, belongs to $\mathscr{L}^{q_2}$.

The other direction is more interesting, both from the mathematical and from the applications point of view. Let $\mathfrak{B}_2 \in \mathscr{L}^{q_2}$. By Proposition A there exists a $R_2(s, s^{-1}) \in \mathbb{R}^{q_2 \times q_2}[s, s^{-1}]$ such that $\mathfrak{B}_2 = \ker R_2(\sigma, \sigma^{-1})$. Now consider $\mathfrak{B}_1 = R(\sigma, \sigma^{-1})\mathfrak{B}_2$. Then $\mathfrak{B}_1 = \{w_1 : \mathbb{Z} \to \mathbb{R}^{q_1} \mid \exists w_2 : \mathbb{Z} \to \mathbb{R}^{q_2}$ such that $R_2(\sigma, \sigma^{-1}) w_2 = 0$ and $w_1 = R(\sigma, \sigma^{-1}) w_2\}$. We will show, more generally, that if $\mathfrak{B}$ is defined in terms of R' and R'' by $\mathfrak{B} := \{w : \mathbb{Z} \to \mathbb{R}^q \mid \exists a : \mathbb{Z} \to \mathbb{R}^q$, such that $R'(\sigma, \sigma^{-1}) w = R''(\sigma, \sigma^{-1}) a\}$, then $\mathfrak{B} \in \mathscr{L}^q$. In order to see that this implies $\mathfrak{B}_1 \in \mathscr{L}^{q_1}$, take

$$R' = \begin{bmatrix} I \\ 0 \end{bmatrix} \quad \text{and} \quad R'' = \begin{bmatrix} R \\ R_2 \end{bmatrix}.$$

Observe that if $U(s, s^{-1})$ and $V(s, s^{-1})$ are unimodular, then $\mathfrak{B} = \{w \mid \exists a$ such that $U(\sigma, \sigma^{-1}) R'(\sigma, \sigma^{-1}) w = U(\sigma, \sigma^{-1}) R''(\sigma, \sigma^{-1}) V(\sigma, \sigma^{-1}) a\}$. Now use the Smith form: choose U and V such that

$$UR''V = \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix}$$

with $\Lambda(s, s^{-1}) := \text{diag}\,[d_1(s, s^{-1}), d_2(s, s^{-1}, \ldots, d_r(s, s^{-1})]$. This yields, with R'_1 and R'_2 defined in the obvious way such that

$$UR' = \begin{bmatrix} R'_1 \\ R'_2 \end{bmatrix},$$

$\mathfrak{B} = \{w \mid \exists a$ such that $R'_1(\sigma, \sigma^{-1})w = \Lambda(\sigma, \sigma^{-1})a$ and $R'_2(\sigma, \sigma^{-1})w = 0\}$. It follows from the lemma that $\text{im}\,\Lambda(\sigma, \sigma^{-1}) = (\mathbb{R}^q)^{\mathbb{Z}}$. Hence $\mathfrak{B} = \ker R'_2(\sigma, \sigma^{-1})$ which shows, using Proposition 4.1A, that $\mathfrak{B} \in \mathscr{L}^q$. □

Proposition 4.1C shows that linear shift invariant closed subspaces of $\mathbb{L}^q$ and polynomial operators in the shift behave in many ways as finite-dimensional vector spaces and matrices.

4.2 (AR), (ARMA), and (MA) systems

4.2.1 *Completeness and its relation to $\mathscr{L}^q$*

Recall that we have called a dynamical system $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ *linear* if $\mathfrak{B}$ is a linear subspace of $(\mathbb{R}^q)^{\mathbb{Z}}$, *time invariant* if $\sigma\mathfrak{B} = \mathfrak{B}$, and *complete* if $\{w \in \mathfrak{B}\} \Leftrightarrow \{w|_{[t_0, t_1]} \in \mathfrak{B}|_{[t_0, t_1]}$ for all $t_0, t_1 \in \mathbb{Z}$ (finite, to be sure)$\}$. Our first result provides a neat mathematical characterization of the linear time invariant complete systems.

PROPOSITION 4.2 $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ is a linear time invariant complete dynamical system if and only if $\mathfrak{B} \in \mathscr{L}^q$.

Proof We only need to worry about the implication $\{\Sigma \text{ complete}\} \Leftrightarrow \{\mathfrak{B} \text{ closed}\}$. To prove ($\Rightarrow$), assume that Σ is complete. Take a sequence $w_n \in \mathfrak{B}$, $n \in \mathbb{Z}_+$, such that $w_n \xrightarrow[n \to \infty]{} w$. We need to show that $w \in \mathfrak{B}$. Since $\mathfrak{B}|_{[t_0, t_1]}$ is a finite-dimensional linear subspace of $\mathbb{R}^{q(t_1 - t_0 + 1)}$, it is closed. Hence, $\{w_n|_{[t_0, t_1]} \in \mathfrak{B}|_{[t_0, t_1]}\} \Rightarrow \{\lim_{n \to \infty} w_n|_{[t_0, t_1]} = w|_{[t_0, t_1]} \in \mathfrak{B}|_{[t_0, t_1]}\}$. By completeness this yields $w \in \mathfrak{B}$. Hence $\mathfrak{B}$ is closed. To prove ($\Leftarrow$), assume that $\mathfrak{B}$ is closed and let $w|_{[t_0, t_1]} \in \mathfrak{B}|_{[t_0, t_1]}$ for all $t_0 \leq t_1$. We need to show that $w \in \mathfrak{B}$. Now, there exists $\tilde{w}^{t_0, t_1} : \mathbb{Z} \to \mathbb{R}^q$ such that $w^{t_0, t_1} := \tilde{w}^{t_0, t_1} \wedge_{t_0^-} w \wedge_{t_1^+} \tilde{w}^{t_0, t_1} \in \mathfrak{B}$. Clearly $w^{t_0, t_1} \xrightarrow[\substack{t_0 \to -\infty \\ t_1 \to +\infty}]{} w$. Hence $w \in \mathfrak{B}$, as claimed. □

4.2.2 (*AR*) *systems*

Proposition 4.1A implies that $\{\mathfrak{B}\in\mathcal{L}^q\}\Leftrightarrow\{\mathfrak{B}=\ker R(\sigma,\sigma^{-1})\}$ for some polynomial operator in the shift $R(\sigma,\sigma^{-1})$, while Proposition 4.2 implies that $\{\mathfrak{B}\in\mathcal{L}^q\}\Leftrightarrow\{\Sigma=(\mathbb{Z},\mathbb{R}^q,\mathfrak{B})$ is linear, shift invariant, and complete$\}$. Hence for the class of systems under consideration we can conclude that completeness is equivalent to finite memory. These results allow us to conclude that the linear time invariant complete systems are *precisely* those which are governed by a set of behavioural equations of the form

$$R_L w(t+L)+R_{L-1}w(t+L-1)+\cdots+R_{l+1}w(t+l+1)+R_l w(t+l)=0 \qquad t\in\mathbb{Z}$$

with $R_L, R_{L-1},\ldots,R_{l+1},R_l\in\mathbb{R}^{g\times q}$. These behavioural equations consist of g scalar linear equations involving as parameters the entries of the matrices $R_L,\ldots,R_l$ and as variables the signal components $w_1, w_2,\ldots,w_q$ and their lags. Essentially all models used in econometrics, in signal processing, and in discrete time linear control are indeed of this type.

We will write these equations compactly in the convenient polynomial form

$$R(\sigma,\sigma^{-1})w=0 \tag{AR}$$

with $R(s,s^{-1})=R_Ls^L+R_{L-1}s^{L-1}+\cdots+R_{l+1}s^{l+1}+R_ls^l\in\mathbb{R}[s,s^{-1}]$, and refer to them as (*AR*) (*AutoRegressive*) *systems.*

The polynomial matrix R determines $\mathfrak{B}$ uniquely, but the converse is not true. In fact, if $U(s,s^{-1})\in\mathbb{R}^{g\times g}[s,s^{-1}]$ is unimodular then $\ker R(\sigma,\sigma^{-1})$ will be equal to $\ker U(\sigma,\sigma^{-1})R(\sigma,\sigma^{-1})$. This implies (cfr. the Smith form) that $\mathfrak{B}\in\mathcal{L}^q$ can always be represented as $\ker R(\sigma,\sigma^{-1})$ with $R(s,s^{-1})\in\mathbb{R}^{g\times q}[s,s^{-1}]$ having $\text{rank}_{\mathbb{R}(s)}R(s,s^{-1})=g$, i.e. with R of full row rank.

Consider $\Sigma=(\mathbb{Z},\mathbb{R}^q,\mathfrak{B})$ with $\mathfrak{B}\in\mathcal{L}^q$. A vector polynomial $p(s,s^{-1})\in\mathbb{R}^{1\times q}[s,s^{-1}]$ is called an *annihilator* of $\mathfrak{B}$ if $p(\sigma,\sigma^{-1})\mathfrak{B}=0$ i.e., if $\{w\in\mathfrak{B}\}\Rightarrow\{p(\sigma,\sigma^{-1})w=0\}$. Denote the set of all annihilators of $\mathfrak{B}$ by $\mathfrak{B}^\perp$. The basic fact leading to Lemma 4.1(i) is that $\mathfrak{B}^\perp$ is a submodule of $\mathbb{R}^{1\times q}[s,s^{-1}]$. If $\mathfrak{B}=\ker R(\sigma,\sigma^{-1})$ with $R(s,s^{-1})\in\mathbb{R}^{g\times q}[s,s^{-1}]$ then, in the obvious notation, $\mathfrak{B}^\perp=\mathbb{R}^{1\times q}[s,s^{-1}]R(s,s^{-1})$. By identifying elements of $\mathbb{R}^{1\times q}[s,s^{-1}]$ with elements of $(\mathbb{R}^q)^{\mathbb{Z}}$, it is possible to view $\mathfrak{B}^\perp$ as defining a linear time invariant system $\Sigma'=(\mathbb{Z},\mathbb{R}^q,\mathfrak{B}^\perp)$, called the *dual* of Σ. Since all elements of $\mathfrak{B}^\perp$ have compact support, Σ' is, of course, not complete. We will not pursue its properties any further in this paper.

4.2.3 (*ARMA*) *systems*

As we have argued in section 1, the natural starting point for what we have called *models from first principles* are models which involve latent variables. Within the context of linear time invariant complete systems (with signal spaces and latent variable spaces which are *finite*-dimensional), this means that we should expect to start with model equations of the form

$$R_1(\sigma, \sigma^{-1})\boldsymbol{w} = R_2(\sigma, \sigma^{-1})\boldsymbol{a} \qquad \text{(ARMA)}$$

where $\boldsymbol{w}: \mathbb{Z} \to \mathbb{R}^q$ and $\boldsymbol{a}: \mathbb{Z} \to \mathbb{R}^d$ denote the signal and latent variables and the entries of the coefficient matrices of $R_1(s, s^{-1}) \in \mathbb{R}^{f \times q}[s, s^{-1}]$, $R_2(s, s^{-1}) \in \mathbb{R}^{f \times d}[s, s^{-1}]$ denote the parameters of the model; d equals the number of latent variables involved in our model, and f denotes the number of laws relating the variables. We will call the above model an *(ARMA)* (*AutoRegressive Moving Average) system.*

The external behaviour of this (ARMA) system is given by $\mathfrak{B} = (R_1(\sigma, \sigma^{-1}))^{-1}(\operatorname{im} R_2(\sigma, \sigma^{-1}))$. Proposition 4.1C implies that the latent variables can always be completely eliminated, i.e., *there will exist a* g, $0 \leq g \leq q$, *and a polynomial matrix* $R(s, s^{-1}) \in \mathbb{R}^{g \times q}[s, s^{-1}]$ *such that the (AR) model* $R(\sigma, \sigma^{-1})\boldsymbol{w} = 0$ *defines exactly the same behaviour as the external w-behaviour of the original (ARMA) model.*

A completely analogous result about elimination of latent variables holds for linear time invariant differential equations (we have seen an example of this in section 1.1.2). These results are fully analogous to the representation result obtained in Theorem 2.5 for discrete event systems. It is matter of conjecture that the behaviour or certain (smooth) classes of differential equations $\boldsymbol{x} = f \circ(\boldsymbol{x}, \boldsymbol{u}); \boldsymbol{y} = r \circ(\boldsymbol{x}, \boldsymbol{u}); \boldsymbol{w} = (\boldsymbol{u}, \boldsymbol{y})$ can be written as the behaviour of a high order differential equation $F \circ(\boldsymbol{w}, \boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(l)}) = 0$. More precisely, we conjecture that if $f: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $r: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ are C^∞-functions and if (in the notation of section 2) $\simeq = \underset{\sim}{+}$, then there will exist an integer $l \in \mathbb{Z}_+$ and a map $F: (\mathbb{R}^m \times \mathbb{R}^p)^{l+1} \to \{0, 1\}$ such that the set

$$\{(\boldsymbol{u}, \boldsymbol{y}) \in C^\infty(\mathbb{R}; \mathbb{R}^m \times \mathbb{R}^p) \mid \exists \boldsymbol{x} \in \mathscr{C}^\infty(\mathbb{R}; \mathbb{R}^n) \quad \text{such that} \quad \dot{\boldsymbol{x}} = f \circ(\boldsymbol{x}, \boldsymbol{u}); \boldsymbol{y} = r \circ(\boldsymbol{x}, \boldsymbol{u})\}$$

will be equal to

$$\{(\boldsymbol{u}, \boldsymbol{y}) \in \mathscr{C}^\infty(\mathbb{R}; \mathbb{R}^m \times \mathbb{R}^p) \mid F(\circ(\boldsymbol{u}, \boldsymbol{y}), (\boldsymbol{u}^{(1)}, \boldsymbol{y}^{(1)}), \ldots, (\boldsymbol{u}^{(l)}, \boldsymbol{y}^{(l)})) = 0\}.$$

This conjecture is valid, as we have just seen, if f and r are linear.

An area of application where these results are relevant is *failure detection*. Assume that the attributes of a system in failure free operation obey a number of linear time invariant laws, an (AR) system of equations. Assume that we want to investigate, by monitoring a subset of the attributes, whether or not the system is failure free. By considering the measured attributes as the signals and the remaining attributes as latent variables, we can consider the model as an (ARMA) system. We have seen that the measured attributes must however satisfy an (AR) system of equations on their own. Monitoring these behavioural relations and signalling when the left-hand side exceeds a threshold offers a reasonable method for detecting failures in the original system.

4.2.4 *Model complexity*

The result about the elimination of latent variables has important implications in model building. Proposition 4.1C implies that a linear time invariant complete system can always be described by at most q recursive (AR)-type equations. Model building often proceeds by the reductionist method of *tearing*, in which a system is viewed as an interconnection of subsystems and the modelling proceeds by *zooming in* on the individual subsystems. The attributes describing the subsystems will now be of two kinds: one class will consist of the variables which we actually want to model and a second class will consist of the variables through which the subsystem is interconnected into the overall system. The complete model will consist of a model for each of the subsystems together with a set of relations expressing the interconnection constraints. In this ultimate model, the interconnection variables should logically be regarded as latent variables. Our result of section 4.2.3 tells us that, as far as the number of equations is concerned, the ultimate model will consist of at most q equations, regardless of the number of auxiliary latent variables and subsystems which had to be introduced in the tearing process.

4.2.5 *(MA) systems*

A special class of (ARMA) systems is given by

$$w = M(\sigma, \sigma^{-1})a \tag{MA}$$

where $w \in \mathbb{R}^q$ and $a \in \mathbb{R}^d$ denote the signal and latent variables and $M(s, s^{-1}) \in \mathbb{R}^{q \times d}[s, s^{-1}]$ the parameters of the model. Clearly (MA) does not constrain the latent variables a and hence its external w-behaviour is simply given by $\mathfrak{B} = \operatorname{im} M(\sigma, \sigma^{-1})$. We will call such a model an *(MA) (Moving Average) system.*

It turns out that (MA) systems have a very neat system theoretic characterization: they are the controllable linear time invariant complete systems.

4.3 Controllability and observability

4.3.1 *(MA) systems are controllable*

Recall from section 1.4.5 that we have called the dynamical system $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ *controllable* if for all $w_1^- \in \mathfrak{B}|_{(-\infty,0)}$ and $w_2^{0+} \in \mathfrak{B}|_{[0,\infty)}$ there exists a $t \geq 0$ and $w:[0,t) \to \mathbb{R}^q$ such that $w_1^- \wedge_0 w \wedge_t \sigma^t w_2^{0+} \in \mathfrak{B}$. The following proposition identifies controllability with (MA) systems and connects it with a special class of (AR) representations

PROPOSITION 4.3 Let $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ be a dynamical system with $\mathfrak{B} \in \mathscr{L}^q$. Then the following conditions are equivalent:

(i) Σ is controllable;
(ii) $\mathfrak{B} = \mathfrak{B}^{\text{compact/closure}}$. Here $\mathfrak{B}^{\text{compact}}$ denotes the elements of $\mathfrak{B}$ with compact support, i.e., $\mathfrak{B}^{\text{compact}} := \{w \in \mathfrak{B} \mid \exists t_{-1}, t_1 \text{ such that } w(t) = 0 \text{ for } t \notin [t_{-1}, t_1]\}$, and the closure is, of course, with respect to the topology of pointwise convergence;
(iii) $\exists m \in \mathbb{Z}_+$ and $M(s, s^{-1}) \in \mathbb{R}^{q \times m}[s, s^{-1}]$ such that $\mathfrak{B} = \operatorname{im} M(\sigma, \sigma^{-1})$;

(iv) if $R(s,s^{-1})\in\mathbb{R}^{g\times q}[s,s^{-1}]$ is such that $\mathfrak{B}=\ker R(\sigma,\sigma^{-1})$, then $\mathrm{rank}_{\mathbb{C}} R(\lambda,\lambda^{-1})$ will be constant for $0\neq\lambda\in\mathbb{C}$.

Proof (Outline): The equivalence (ii)⇔(iii) follows from Proposition 4.1C. To prove the equivalence of (i), (ii), and (iv), first verify that if any of the conditions hold for $\mathfrak{B}$ and if $V(s,s^{-1})\in\mathbb{R}^{q\times q}[s,s^{-1}]$ is unimodular then the corresponding condition also holds for $V(\sigma,\sigma^{-1})\mathfrak{B}$. Now let $\mathfrak{B}=\ker R(\sigma,\sigma^{-1})$ with R of full row rank. Then there exist unimodular matrices $U(s,s^{-1})\in\mathbb{R}^{g\times g}[s,s^{-1}]$ and $V(s,s^{-1})\in\mathbb{R}^{q\times q}[s,s^{-1}]$ such that $URV=[\Lambda\,\vdots\,0]$ with $\Lambda=\mathrm{diag}[d_1,d_2,\ldots,d_g]$ and $0\neq d_i$. Obviously $\mathfrak{B}=V(\sigma,\sigma^{-1})\ker[\Lambda(\sigma,\sigma^{-1})\,\vdots\,0]$. Finally prove that $\ker[\Lambda(\sigma,\sigma^{-1})\,\vdots\,0]$ satisfies any of the conditions (i) to (iv) of Proposition 4.3 if and only if all the d_i's are unimodular. ■

Consider $\Sigma=(\mathbb{Z},\mathbb{R}^q,\mathfrak{B})$ with $\mathfrak{B}\in\mathscr{L}^q$. Define the *controllable part* of Σ as $\Sigma^{\text{controllable}}:=(\mathbb{Z},\mathbb{R}^q,\mathfrak{B}^{\text{controllable}})$ with $\mathfrak{B}^{\text{controllable}}:=\mathfrak{B}^{\text{compact/closure}}$ (see section 4.1.3). Obviously $\Sigma^{\text{controllable}}$ defines a subsystem of Σ ($\Sigma'=(\mathbb{Z},\mathbb{R}^q,\mathfrak{B}')$ is called a *subsystem* of Σ if $\mathfrak{B}'\subseteq\mathfrak{B}$); $\Sigma^{\text{controllable}}$ is in fact the largest controllable subsystem of Σ. It is easy to compute $\mathfrak{B}^{\text{controllable}}$ starting from an (AR) representation of $\mathfrak{B}=\ker R(\sigma,\sigma^{-1})$. Indeed (cfr. the Smith form) $R(s,s^{-1})\in\mathbb{R}^{g\times q}[s,s^{-1}]$ may be written as $R(s,s^{-1})=F(s,s^{-1})R'(s,s^{-1})$ with $F(s,s^{-1})\in\mathbb{R}^{g\times g}[s,s^{-1}]$ having $\det F\neq 0$ and $R'(s,s^{-1})\in\mathbb{R}^{g\times q}[s,s^{-1}]$ having $\mathrm{rank}_{\mathbb{C}} R'(\lambda,\lambda^{-1})$ constant for all $0\neq\lambda\in\mathbb{C}$. Then $\mathfrak{B}^{\text{controllable}}=\ker R'(\sigma,\sigma^{-1})$.

4.3.2 *Observable (ARMA) representations*

Consider the (ARMA) system

$$R(\sigma,\sigma^{-1})w=M(\sigma,\sigma^{-1})a$$

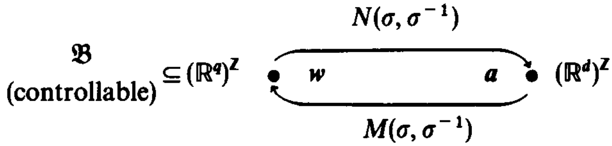
Let $\Sigma_a=\{\mathbb{Z},\mathbb{R}^q,\mathbb{R}^d,\mathfrak{B}_a\}$ with $\mathfrak{B}_a=\ker[R(\sigma,\sigma^{-1})\,\vdots\,-M(\sigma,\sigma^{-1})]$ the extended behaviour. Following section 2.1.1, we will call this (ARMA) system *observable* if $\{(w,a'),(w,a'')\in\mathfrak{B}_a\}\Rightarrow\{a'=a''\}$. It is easy to see that this is equivalent to requiring $\{(0,a)\in\mathfrak{B}_a\}\Leftrightarrow\{a=0\}$.

Proposition 4.4 An (ARMA) system is observable iff $\mathrm{rank}_{\mathbb{C}} M(\lambda,\lambda^{-1})=d$ for all $0\neq\lambda\in\mathbb{C}$.

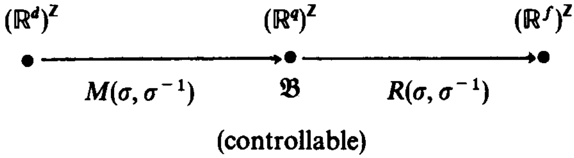
Proof Follows immediately from Lemma 4.1 ■

The above proposition implies that in an (MA) representation a will be observable from w iff rank $M(\lambda,\lambda^{-1})=d$ for all $0\neq\lambda\in\mathbb{C}$. Such observable (MA) representations of reachable systems with $\mathfrak{B}\in\mathscr{L}^q$ always exist. It follows that the latent variable a will then be related to w by $a=N(\sigma,\sigma^{-1})w$ for some

$N(s, s^{-1}) \in \mathbb{R}^{d \times q}[s, s^{-1}]$. Clearly $NM = I$. Schematically we have



It can be shown that we can in fact take M and N to be polynomial matrices in s (or s^{-1}). The relation between (AR) and (MA) representations of reachable systems can be schematically expressed by the diagram below.



This diagram is a short exact sequence with $\operatorname{im} M(\sigma, \sigma^{-1}) = \ker R(\sigma, \sigma^{-1}) = \mathfrak{B}$.

4.3.3 *Relation with the Hautus test*

Consider the classical state space system $\sigma x = Ax + Bu; y = Cx + Du$. Here $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times m}$, and $D \in \mathbb{R}^{p \times m}$.

First, examine the controllability of $\sigma x = Ax + Bu$, viewed as defining the system $\Sigma_{A,B} = (\mathbb{Z}, \mathbb{R}^n \times \mathbb{R}^m, \mathfrak{B})$ with $\mathfrak{B} = \ker [I\sigma - A \,\vdots\, -B]$. $\Sigma_{A,B}$ defines an (AR) system in the signal variable

$$\begin{bmatrix} x \\ \cdots \\ u \end{bmatrix}.$$

According to Proposition 4.3, this system is controllable iff $\operatorname{rank}_{\mathbb{C}} [I\lambda - A \,\vdots\, B] = n$ for all $0 \neq \lambda \in \mathbb{C}$. By Proposition 2.7, this is also equivalent to state point controllability.

Next, examine the observability of $\sigma x = Ax + Bu; y = Cx + Du$ viewed as an (ARMA) system with

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

the signal variable and x the latent variable. According to Proposition 4.4, this system is observable iff

$$\operatorname{rank}_{\mathbb{C}} \begin{bmatrix} I\lambda - A \\ \cdots \\ C \end{bmatrix} = n \qquad \text{for all } 0 \neq \lambda \in \mathbb{C}.$$

The conditions: $\operatorname{rank}_{\mathbb{C}}[I\lambda - A \vdots B] = n$ for controllability and

$$\operatorname{rank}_{\mathbb{C}}\begin{bmatrix} I\lambda - A \\ \cdots\cdots \\ C \end{bmatrix} = n$$

for observability are known as the *Hautus test* [21] (actually, Hautus requires these rank conditions to hold for all $\lambda \in \mathbb{C}$—and not just for $\lambda \neq 0$—this minor difference is due to the fact that we are working with time axis $T = \mathbb{Z}$ instead of $T = \mathbb{Z}_+$). In any case, our condition for controllability: $\operatorname{rank}_{\mathbb{C}} R(\lambda, \lambda^{-1})$ constant for $0 \neq \lambda \in \mathbb{C}$, and for observability: $\operatorname{rank}_{\mathbb{C}} M(\lambda, \lambda^{-1})$ constant for $0 \neq \lambda \in \mathbb{C}$ can be viewed as considerable generalizations of these results.

4.4 Autonomous systems

4.4.1 *Representations of autonomous systems*

Recall that a dynamical system $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ is called *autonomous* if there exists a map $f: \mathfrak{B}^- \to \mathfrak{B}^{0+}$ such that $\{w^- \wedge_0 w^{0+} \in \mathfrak{B}\} \Leftrightarrow \{w^{0+} = f(w^-)\}$, in other words if the past implies the future. These are numerous equivalent characterizations of autonomous linear time invariant complete systems. We collect a few in the following proposition.

PROPOSITION 4.5 Let $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ be a dynamical system with $\mathfrak{B} \in \mathscr{L}^q$. Then the following conditions are equivalent:

(i) Σ is autonomous;
(ii) $\mathfrak{B}$ is finite dimensional;
(iii) $\exists R(s, s^{-1}) \in \mathbb{R}^{q \times q}[s, s^{-1}]$ with $\det R \neq 0$ such that $\mathfrak{B} = \ker R(\sigma, \sigma^{-1})$;
(iv) $\exists t \in \mathbb{Z}_+$ and a linear map $f: \mathfrak{B}|_{[0,t)} \to (\mathbb{R}^q)^{\mathbb{Z}}$ such that $\{w \in \mathfrak{B}\} \Leftrightarrow \{w = f(w|_{[0,t)})\}$.

Proof Our plan is to run the circle (ii)$\Rightarrow$(iv)$\Rightarrow$(i)$\Rightarrow$(iii)$\Rightarrow$(ii). (ii)$\Rightarrow$(iv): Define $\mathfrak{B}_t := \mathfrak{B}|_{[0,t]}$. Clearly $\lim_{t\to\infty} \dim \mathfrak{B}_t = \dim \mathfrak{B}$. Hence, if (ii) holds, $\exists t' \in \mathbb{Z}_+$ such that $\dim \mathfrak{B}_{t'} = \dim \mathfrak{B}$. Now the projection $\pi_{t'}: \mathfrak{B} \to \mathfrak{B}_{t'}$ defined by $\pi_{t'} w = w|_{[0,t']}$ is a surjection—hence, since its domain and codomain have the same finite dimension, a bijection. This yields (iv). (iv)$\Rightarrow$(i) is obvious. (i)$\Rightarrow$(iii): Let $\mathfrak{B} = \ker R(\sigma, \sigma^{-1})$, and assume that R is of full now rank over $\mathbb{R}(s)$. Let $U(s, s^{-1})$ and $V(s, s^{-1})$ be unimodular matrices such that $R = U[\Lambda \vdots 0]V$ with $\Lambda = \operatorname{diag}[d_1, d_2, \ldots, d_g]$ and $d_i \neq 0$. Let $\mathfrak{B}' := V(\sigma, \sigma^{-1})\mathfrak{B}$. Observe that $\{\mathfrak{B}$ autonomous$\} \Leftrightarrow \{\{0^- \wedge_0 w^{0+} \in \mathfrak{B}\} \Leftrightarrow \{w^{0+} = 0^{0+}\}\}$. Now deduce from this and the fact that $V(\sigma, \sigma^-)$ and $V^{-1}(\sigma, \sigma^{-1})$ are polynomial operators in the shift, that $\{\mathfrak{B}$ autonomous$\} \Leftrightarrow \{\mathfrak{B}'$ autonomous$\}$. Clearly $\mathfrak{B}' = \ker[\Lambda(\sigma, \sigma^{-1}) \vdots 0]$ is autonomous iff $g = q$, i.e., Λ, and hence $R = U\Lambda V$, must have rank q over $\mathbb{R}(s)$. Hence (i)$\Rightarrow$(iii). To see the implication (iii)$\Rightarrow$(ii), observe that if $\Lambda = \operatorname{diag}[d_1, d_2, \ldots, d_q]$, then $\ker \Lambda(\sigma,\sigma^{-1}) = e_1 \ker d_1(\sigma,\sigma^{-1}) \oplus e_2 \ker d_2(\sigma,\sigma^{-1}) \oplus \cdots \oplus e_q \ker d_q(\sigma,\sigma^{-1})$ with e_i as in the proof of Lemma 4.1.3 and with $\ker d_i(\sigma,\sigma^{-1})$ viewed as a subspace of $(\mathbb{R})^{\mathbb{Z}}$. Now examine $\dim \ker d(\sigma,\sigma^{-1})$ for $d(s,s^{-1}) = r_L s^L + \cdots + r_l s^l$ and $r_L e_l \neq 0$. Then (see the proof of Lemma 4.1) $\dim \ker d(\sigma,\sigma^{-1}) = L - l$. ■

It follows immediately from this proof that if $R(s,s^{-1}) \in \mathbb{R}^{q\times q}[s,s^{-1}]$ has $0 \neq \det R(s,s^{-1}) = \alpha_L s^L + \cdots + \alpha_l s^l$, with $\alpha_L \alpha_l \neq 0$, then $\mathfrak{B} = \ker R(\sigma,\sigma^{-1})$ has $\dim \mathfrak{B} = L - l$.

4.5 Inputs and outputs

4.5.1 *Structured (AR) systems*

For a dynamical system in the class studied in this section, i.e., $\Sigma = (\mathbb{Z}, W_1 \times W_2, \mathfrak{B})$ with $W_1 = \mathbb{R}^{q}$, $W_2 = \mathbb{R}^{q_2}$, and $\mathfrak{B} \in \mathscr{L}^{q_1+q_2}$ it is possible to express the structural properties introduced in section 3 and relating the signal variables w_1 and w_2 as being induced by special types of (AR) relations. From Proposition 4.1 we know that there will always exist polynomial matrices R_1, R_2 such that the behavioural equations are given in (AR) form by

$$R_1(\sigma,\sigma^{-1})w_1 = R_2(\sigma,\sigma^{-1})w_2$$

Observability: w_2 will be observable from w_1 iff $\mathfrak{B}$ admits an (AR) representation of the following form

$$R_1'(\sigma,\sigma^{-1})w_1 = 0; \quad w_2 = R_1''(\sigma,\sigma^{-1})w_1.$$

For past observability we will be able to choose R_1'' such that it contains only negative powers of s and for future observability such that it contains only non-negative powers.

Processing requires a slightly less restrictive (AR) relation: w_2 processes w_1 iff $\mathfrak{B}$ admits an (AR) relation of the following form:

$$R_1'(\sigma,\sigma^{-1})w_1 = 0; \quad R_2''(\sigma,\sigma^{-1})w_2 = R_1''(\sigma,\sigma^{-1})w_1$$

with $R_s''(s,s^{-1}) \in \mathbb{R}^{q_2\times q_2}[s,s^{-1}]$, $\det R_2'' \neq 0$. By premultiplication with an unimodular matrix, if necessary, this second relation can actually be written as

$$w_2(t) = \sum_{t'=1}^{L} R_{t'}''\, w_2(t-t') + \sum_{t'=l}^{L} R_{t'}' w_1(t-t')$$

Free: w_1 will be a set of free variables ($\Leftrightarrow$ {locally free} $\Leftrightarrow$ {$\mathfrak{B}_1 := P_{w_1}\mathfrak{B}$ equals $(\mathbb{R}^{q_1})^{\mathbb{Z}}$}) iff $\mathfrak{B}$ admits an (AR) representation with R_1, R_2 satisfying

$$\mathrm{im}_{\mathbb{R}(s)} R_1(s,s^{-1}) \subseteq \mathrm{im}_{\mathbb{R}(s)} R_2(s,s^{-1})$$

or equivalently, since we can always take $[R_1 \vdots -R_2]$ to have full row rank, with an R_2 of full row rank.

We will not prove these claims explicitly, but concentrate on the following proposition which treats the i/o case. Recall that in an i/o system the output variable processes the input variable, which itself must be locally free.

PROPOSITION 4.6 Let $\Sigma = (\mathbb{Z}, W_1 \times W_2, \mathfrak{B})$ be a dynamical system with $W_1 = \mathbb{R}^{q_1}$, $W_2 = \mathbb{R}^{q_2}$, and $\mathfrak{B} \in \mathscr{L}^{q_1+q_2}$. Then Σ defines an i/o system with w_1 as input and w_2 as output iff $\mathfrak{B}$ admits a representation as

$$R_2(\sigma, \sigma^{-1})w_2 = R_1(\sigma, \sigma^{-1})w_1$$

with $R_1(s, s^{-1}) \in \mathbb{R}^{q_2 \times q_1}[s, s^{-1}]$, $R_2(s, s^{-1}) \in \mathbb{R}^{q_2 \times q_2}[s, s^{-1}]$, and $\det R_2 \neq 0$. It defines a non-anticipating i/o system if, in addition, $R_2^{-1}(s, s^{-1})R_1(s, s^{-1}) \in \mathbb{R}_+^{q_2 \times q_1}(s)$.

Proof $\mathfrak{B}$ admits an (AR) representation with $[R_1 \vdots -R_2]$ of full row rank. We will first prove that w_1 will be free iff R_2 is of full row rank. That this is sufficient follows from Lemma 4.1. That it is also necessary may be seen as follows. If R_2 is not of full row rank then there exists $f(s, s^{-1}) \in \mathbb{R}^{1 \times g}[s, s^{-1}]$ such that $fR_2 = 0$ and $fR_1 \neq 0$. This yields $\{(w_1, w_2) \in \mathfrak{B}\} \Rightarrow \{f(\sigma, \sigma^{-1})R_1(\sigma, \sigma^{-1})w_1 = 0\}$, contradicting that w_1 is free. Next, we will prove that w_2 processes w_1 iff $\ker^{\mathbb{R}(s)} R_2(s, s^{-1}) = 0$. Now $\{(0, w_2) \in \mathfrak{B}\} \Leftrightarrow \{w_2 \in \ker R_2(\sigma, \sigma^{-1})\}$ and processing is easily seen to be equivalent to requiring $\{(0, w_2) \in \mathfrak{B}$ and $w_2^- = \mathbf{0}\} \Leftrightarrow \{w_2 = 0\}$. In other words, $\ker R_2(\sigma, \sigma^{-1})$ must define the behaviour of an autonomous system. Now apply Proposition 4.5. This yields the first part of the proposition. To see the non-anticipation condition, observe that if R_2 is square and $\det R_2 \neq 0$, we can express $w_2(t)$ in terms of w_1 and $w_2(t-1), \ldots, w_2(t-L)$, as given in the processing section. Now use this expression to verify that $\{l \geq 0\} \Leftrightarrow \{\{w_1 \in \mathfrak{B}_1, w_1(t) = 0 \text{ for } t \leq 0\}\} \Rightarrow \{\exists w_2 \in \mathfrak{B}_2 \text{ with } w_2(t) = 0 \text{ for } t \leq 0 \text{ such that } \{(w_1, w_2) \in \mathfrak{B}\}\}$. ∎

4.5.2 *Inputs and outputs always exist!*

THEOREM 4.1 *Consider $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ with $\mathfrak{B} \in \mathscr{L}^q$. Then there exists a component-wise partition of $\mathbb{R}^q$ into $\mathbb{R}^q = \mathbb{R}^m \times \mathbb{R}^p$ such that $\Sigma = (\mathbb{Z}, \mathbb{R}^{m+p}, \mathfrak{B})$ is a non-anticipating input/output structure, in other words, such that $\mathfrak{B}$ is represented by*

$$P(\sigma, \sigma^{-1})y = Q(\sigma, \sigma^{-1})u \quad w = \begin{bmatrix} u \\ \cdots \\ y \end{bmatrix} \tag{i/o}$$

with $P(s, s^{-1}) \in \mathbb{R}^{p \times p}[s, s^{-1}]$, $Q(s, s^{-1}) \in \mathbb{R}^{p \times m}[s, s^{-1}]$, $\det P \neq 0$, *and* $P^{-1}(s, s^{-1})Q(s, s^{-1}) \in \mathbb{R}_+^{p \times m}(s)$.

With a *componentwise partition* of $\mathbb{R}^q$ we mean the following. Let $w = \operatorname{col}[w^1, w^2, \ldots, w^q]$ be a vector in $\mathbb{R}^q$ with $w^i \in \mathbb{R}$. Partition $\boldsymbol{q} := \{1, 2, \ldots, q\}$ into two disjoint sets, $\boldsymbol{q}_1 = \{a_1, a_2, \ldots, a_m\}$ and $\boldsymbol{q}_2 = \{b_1, b_2, \ldots, b_p\}$, such that $\boldsymbol{q}_1 \cap \boldsymbol{q}_2 = \varnothing$ and $\boldsymbol{q}_1 \cup \boldsymbol{q}_2 = \boldsymbol{q}$. Now consider the vectors $w_1 := \mathrm{col}[w^{a_1}, w^{a_2}, \ldots, w^{a_m}]$ and $w_2 := \mathrm{col}[w^{b_1}, w^{b_2}, \ldots, w^{b_p}]$. The partition $w \approx \mathrm{col}[w_1, w_2]$ is what we mean by a componentwise partition of w. Thus Theorem 4.1 states that separation of w into non-anticipating input and output variables is possible by considering the signal not only in the way a mathematician looks upon $\mathbb{R}^q$ as an abstract q-dimensional real vector space, but also in the manner an engineer or an econometrician looks upon a vector: as a finite sequence of real numbers, all having a concrete interpretation.

Proof Let $\mathfrak{B} = \ker R(\sigma, \sigma^{-1})$ with R of full row rank. It is well-known that there exists an unimodular matrix $U(s, s^{-1})$ such that UR is row proper, that is $R' = UR$ is of the form $R'(s, s^{-1}) = R_L s^L + \cdots$ with $R_L \in \mathbb{R}^{g \times q}$ of full row rank. Now let $\pi \in \mathbb{R}^{q \times q}$ be a permutation matrix such that $R_L \pi = [\tilde{R}_1 \vdots \tilde{R}_2]$ has $\det \tilde{R}_2 \neq 0$. Define

$$\begin{bmatrix} w_1 \\ \cdots \\ w_2 \end{bmatrix} = \pi^{-1} w$$

to be a partition of w with $w_1 \in \mathbb{R}^{q-g}$ and $w_2 \in \mathbb{R}^g$ and verify that $R'(s, s^{-1})\pi$ satisfies the conditions required by Proposition 4.6 for it to define a non-anticipating i/o system. ■

4.6 The frequency response and the transfer function

4.6.1 *The frequency response*

In this section it will be convenient to consider complex systems $\Sigma = (\mathbb{Z}, \mathbb{C}^q, \mathfrak{B})$ obtained, if necessary, from $\tilde{\Sigma} = (\mathbb{Z}, \mathbb{R}^q, \tilde{\mathfrak{B}})$ by *complexification*: $\{w \in \mathfrak{B}\} :\Leftrightarrow \{$both the real and complex part of w belong to $\tilde{\mathfrak{B}}\}$.

Let $\Sigma = (\mathbb{Z}, \mathbb{C}^q, \mathfrak{B})$ be a linear time invariant system. Observe that every linear time invariant systems $\Sigma' = (\mathbb{Z}, \mathbb{C}^q, \mathfrak{B}')$ with $\mathfrak{B}'$ *one-dimensional* must necessarily be of the form $\mathfrak{B}' = \{\alpha a f_\rho \mid \alpha \in \mathbb{C}\}$ for some $0 \neq a \in \mathbb{C}^q$ and f_ρ the exponential time series, parametrized by $\rho \in \mathbb{C}$ defined by $f_\rho : \mathbb{Z} \to \mathbb{C}, f_\rho(t) := e^{\rho t}$. Now examine all subsystems $\Sigma' = (\mathbb{Z}, \mathbb{C}^q, \mathfrak{B}')$ of $\Sigma = (\mathbb{Z}, \mathbb{C}^q, \mathfrak{B})$ with $\mathfrak{B}'$ one-dimensional. This associates with Σ a map from $\rho \in \mathbb{C}$ to the set $\mathbb{A}_\rho \subseteq \mathbb{C}^q$ defined by $\{a \in \mathbb{A}_\rho\} :\Leftrightarrow \{a f_\rho \in \mathfrak{B}\}$. It is easily seen that $\mathbb{A}_\rho$ is a linear subspace of $\mathbb{C}^q$. Let $\mathcal{G}(-, \mathbb{C}^q)$ denote the set of all linear subspaces of $\mathbb{C}^q$. We will call the mapping $\mathbb{F} : \mathbb{C} \to \mathcal{G}(-, \mathbb{C}^q)$ defined by $\rho \overset{\mathbb{F}}{\mapsto} \mathbb{A}_\rho$, the *frequency response of* Σ. If $\mathbb{A}_\rho \neq 0$, then we will call ρ a *characteristic frequency* of Σ and $\mathbb{A}_\rho$ the associated *characteristic amplitudes.*

Now consider $\Sigma = (\mathbb{Z}, \mathbb{C}^q, \mathfrak{B})$ with $\mathfrak{B} \in \mathscr{L}^q$ (with $\mathscr{L}^q$ defined over $\mathbb{C}$, of course). Let $\mathfrak{B} = \ker R(\sigma, \sigma^{-1})$ with $R(s, s^{-1}) \in \mathbb{R}^{g \times q}[s, s^{-1}]$. Then the frequency response is given by $\mathbb{F}: \rho \in \mathbb{C} \mapsto \mathbb{A}_\rho = \ker R(e^\rho, e^{-\rho}) \in \mathscr{G}(-, \mathbb{C}^q)$. This implies that there exist a constant, m, $0 \leq m \leq q$, called the *normal dimension* of $\mathbb{A}_\rho$, such that $\dim \mathbb{A}_\rho = m$ for all except possibly a finite number of ρ's, where $\dim \mathbb{A}_\rho > m$. We will call these the *singular points* of the frequency response.

Two special cases are worth noting:

1. *The normal dimension* $m = 0$. This corresponds to the case that the system is autonomous.
2. *The frequency response contains no singular points.* Using Proposition 4.3(iv), we immediately see that this case corresponds to controllable systems. Let $\Sigma = (\mathbb{Z}, \mathbb{C}^q, \mathfrak{B})$, with $\mathfrak{B} \in \mathscr{L}^q$, be controllable. Let $\mathfrak{B} = \operatorname{im} M(\sigma, \sigma^{-1})$ with $M(s, s^{-1}) \in \mathbb{R}^{q \times m}[s, s^{-1}]$. Then the frequency response is given by $\mathbb{F}: \rho \in \mathbb{C} \mapsto \mathbb{A}_\rho = \operatorname{im} M(e^\rho, e^{-\rho})$. Note that if $M(s, s^{-1})$ is full column rank, then the number of latent variables in (MA) equals the normal dimension of the frequency response.

The frequency response is, of course, uniquely defined from the behaviour, but the converse is, unfortunately, not always the case, not even when $\mathfrak{B} \in \mathscr{L}^q$. In order to see this, observe that the systems $\mathfrak{B}_1 = \ker p_1(\sigma)$ with $p_1(s), p_2(s) \in \mathbb{R}[s]$ have already the same frequency response if and only if p_1 and p_2 have the same non-zero roots, disregarding their multiplicity. For controllable systems, however, the frequency response uniquely determines the system. More precisely:

PROPOSITION 4.7 Consider $\Sigma_1(\mathbb{Z}, \mathbb{C}^q, \mathfrak{B}_1)$ and $\Sigma_2 = (\mathbb{Z}, \mathbb{C}^q, \mathfrak{B}_2)$ with $\mathfrak{B}_1, \mathfrak{B}_2 \in \mathscr{L}^q$. Assume that Σ_1 is controllable. Then $\Sigma_1 = \Sigma_2$ if and only if they have the same frequency response.

Proof Use the Smith form for $\mathfrak{B}_1$. ■

Consider the linear time invariant system $\Sigma = (\mathbb{Z}, \mathbb{C}^q, \mathfrak{B})$ with frequency response $\mathbb{A}_\rho$. This defines a (set theoretic) sub-bundle of the trivial bundle $\pi: \mathbb{C} \times \mathbb{C}^q \to \mathbb{C}$, with $\pi(\rho, a) := \rho$, defined by $\{(\rho, a) \in \mathbb{C} \times \mathbb{C}^q \mid a \in \mathbb{A}_\rho\}$. Hence each fibre is a vector space. If $\mathfrak{B} \in \mathscr{L}^q$ then this sub-bundle defines a vector bundle (implying that the fibres $\mathbb{A}_\rho$ all have the same dimension) if and only if $\mathfrak{B}$ is controllable. The study of controllable systems consequently reduces to the study of (algebraic) vector bundles over $\mathbb{C}$ (or over the Riemann sphere for that matter). For general systems $\mathfrak{B} \in \mathscr{L}^q$ one ought to introduce sheaves.

4.6.2 *The transfer function*

Consider the i/o system $\Sigma = (\mathbb{Z}, \mathbb{R}^m \times \mathbb{R}^p, B)$, $\mathfrak{B} \in \mathscr{L}^{m+p}$, with representation (i/o) as in Theorem 4.1. The matrix of rational functions $G(s) \in \mathbb{R}^{p\times m}(s)$ defined by

$$G(s) = (P(s, s^{-1}))^{-1} Q(s, s^{-1}) \qquad \text{(TF)}$$

is called the *transfer function* of (i/o). If $G(s) \in \mathbb{R}_+^{p\times m}(s)$, we will call the transfer function *proper* and if $G(s) \in \mathbb{R}_-^{p\times m}(s)$, we will call it *anti-proper*. Theorem 4.1 allows us to conclude that any $\mathfrak{B} \in \mathscr{L}^q$ always admits a (componentwise) i/o representation with a proper transfer function. Note that we have not incorporated non-anticipation in the notion of input and output, nor is it essential for our definition of transfer function. We will illustrate the convenience of our point of view by means of an example in section 4.6.3.

Consider again the system (i/o). Now look for responses

$$w_i = \begin{bmatrix} u_i \\ \cdots \\ y_i \end{bmatrix} \in \mathfrak{B}, \qquad i = 1, 2, \ldots, m,$$

for which the inputs are the following impulses:

$$u_i(t) = \begin{cases} e_i & \text{for } t = 0 \\ 0 & \text{for } t \neq 0 \end{cases}$$

with $e_i \in \mathbb{R}^m$ the standard basis vectors: $e_i = \operatorname{col}[0, \ldots, 0, 1, 0, \ldots, 0]$ with the 1 in the ith position. Since u is free, we know that such responses w_i indeed exist. In fact, there are many such w_i,s. However, since *y processes u*, there is precisely one for which $w_i(t)$ vanishes for t sufficiently small. We will call this w_i the *right-sided ith channel impulse response* and the mapping $H: \mathbb{Z} \to \mathbb{R}^{p\times m}$ defined by $H(t) := [y_1(t) y_2(t) \cdots y_m(t)]$ the *right-sided impulse response matrix of* the system. The matrices $\ldots, H(-t), \ldots, H(0), H(1), \ldots, H(t), \ldots$ are sometimes called the *Markov parameters* of the system. It follows that to any $u \in (\mathbb{R}^m)^{\mathbb{Z}}$, with $u(t) = 0$ for t sufficiently small, there corresponds exactly one $y \in (\mathbb{R}^p)^{\mathbb{Z}}$ such that

$$\begin{bmatrix} u \\ \cdots \\ y \end{bmatrix} \in \mathfrak{B}$$

and such that $y(t) = 0$ for t sufficiently small. This y is, in fact, given by the convolution

$$y(t) = \sum_{t'=-\infty}^{+\infty} H(t - t') u(t')$$

The transfer function and the right-sided impulse response are related by the z-transform. Indeed, for $0 \neq z \in \mathbb{C}$ and $|z|$ sufficiently small there holds

$$G(z) = \sum_{t=-\infty}^{+\infty} H(t) z^{-t}$$

4.6.3 A smoother

Consider the following simple smoothing algorithm:

$$w_2(t) = \frac{1}{2T+1} \sum_{t'=-T}^{T} w_1(t+t') \tag{Sm}$$

where $T \in \mathbb{Z}_+$. ($2T =$ the *smoothing window*), is a fixed integer. This system is formally defined as $\Sigma = (\mathbb{Z}, \mathbb{R}^2, \mathfrak{B})$ with $\mathfrak{B} = \{w_1, w_2): \mathbb{Z} \to \mathbb{R}^2 \mid (Sm)$ is satisfied for all $t \in \mathbb{Z}\}$. An (AR) polynomial representing this system is

$$R(s, s^{-1}) = [s^T + \cdots s + 1 + s^{-1} + \cdots + s^{-T} \vdots -2T-1)]$$

and an (MA) polynomial for it is

$$M(s, s^{-1}) = \left[\frac{2T+1}{s^T + \cdots + s + 1 + s^{-1} + \cdots + s^{-T}} \right]$$

This system is controllable. It is in input/output form with w_1 the input variable and w_2 the output variable. Its transfer function, which is neither proper nor anti-proper, is

$$\frac{1}{2T+1}(s^T + \cdots + s + 1 + s^{-1} + \cdots + s^{-T})$$

which equals

$$\frac{1}{2T+1} \frac{s^{2T+1} - 1}{s^T(s-1)}$$

However, for the system under consideration, it is possible to reverse the roles of the input and the output. Viewing w_1 as input variable and w_2 as output variable yields the right-sided impulse response.

$$\begin{array}{c} w_1 \rightarrow \\ w_2 \rightarrow \end{array} \left[\begin{array}{ccccccccc} \cdots 0 & 0 & \cdots & 0 & \overset{t=0}{\overset{\downarrow}{1}} & 0 & \cdots & 0 & 0\cdots \\ \cdots 0 & \frac{1}{2T+1} & \cdots & \frac{1}{2T+1} & \frac{1}{2T+1} & \frac{1}{2T+1} & \cdots & \frac{1}{2T+1} & 0\cdots \end{array} \right]$$

while viewing it with w_2 as input variable and w_1 as output variable yields the proper transfer function

$$\frac{2T+1}{s^T + \cdots + s + 1 + s^{-1} + \cdots + s^{-T}}$$

and the right-sided impulse response

$$\begin{array}{c} w_1 \rightarrow \\ w_2 \rightarrow \\ \end{array} \left[\begin{array}{ccccccccccccc} \cdots 0 & \overset{t=0}{\overset{\downarrow}{0}} & 0\cdots 0 & (2T+1) & -(2T+1) & 0\cdots 0 & (2T+1) & -(2T+1) & 0\cdots 0 & (2T+1) & -(2T+1) & 0\cdots \\ \cdots 0 & 1 & \underbrace{0\cdots 0}_{(T-1)\text{zeros}} & 0 & 0 & \underbrace{0\cdots 0}_{(2T-1)\text{zeros}} & 0 & 0 & \underbrace{0\cdots 0}_{(2T-1)\text{zeros}} & 0 & 0 & 0\cdots \end{array} \right]$$

This example illustrates that proper transfer function representations are not always as fundamental as they are made out to be. Indeed, the above system allows an interpretation with *both* a proper and a non-proper transfer function. It is the latter one which has a natural interpretation.

4.6.4 *The transfer function determines the controllable part of a system*

We have seen in section 4.6.1 that the frequency response does not in general determine the behaviour of an uncontrollable system uniquely: some of the fine structure of the behaviour cannot quite be represented by the one-dimensional subsystems defining the frequency response. A similar situation occurs with the transfer function.

Consider $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ with $\mathfrak{B} \in \mathcal{L}^q$. It is easy to see that $\Sigma^{\text{controllable}} = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B}^{\text{controllable}})$ is a controllable linear time invariant complete system. Its frequency response $\mathbb{F}^{\text{controllable}}$ is related to the frequency response $\mathbb{F}$ of Σ as follows. Let $\mathbb{F}: \rho \mapsto \mathbb{A}_\rho$ and m be the normal dimension of $\mathbb{A}_\rho$. Denote by $\mathcal{G}(m, \mathbb{C}^q)$ the set of all m-dimensional subspaces of $\mathbb{C}^q$; $\mathcal{G}(m, \mathbb{C}^q)$ is a compact manifold, called a *Grassmann manifold.* Define $\mathbb{F}^{\text{reg}}$, the *regularization* of $\mathbb{F}$, as the map from $\mathbb{C}$ to $\mathcal{G}(m, \mathbb{C}^q)$, as $\mathbb{F}^{\text{reg}}(\rho) := \lim_{\substack{\rho' \to \rho \\ \rho' \neq \rho}} \mathbb{F}(\rho')$. This limit always exists. Of course, $\mathbb{F}^{\text{reg}}(\rho) = \mathbb{F}(\rho)$ for the ρ's where $\mathbb{F}$ is not singular, but at singular points ρ of $\mathbb{F}$, $\mathbb{F}^{\text{reg}}(\rho)$ will be strictly contained in $\mathbb{F}(\rho)$. We state without proof:

PROPOSITION 4.8 Let $\Sigma = (\mathbb{Z}, \mathbb{C}^q, \mathfrak{B})$ with $\mathfrak{B} \in \mathcal{L}^q$ have frequency response $\mathbb{F}$. Then $\Sigma^{\text{controllable}} := (\mathbb{Z}, \mathbb{C}^q, \mathfrak{B}^{\text{controllable}})$ has frequency response $\mathbb{F}^{\text{reg}}$.

It immediately follows that Σ is controllable iff $\mathbb{F} = \mathbb{F}^{\text{reg}}$.

Now consider two (i/o) systems $\Sigma_1 = (\mathbb{Z}, \mathbb{R}^m \times \mathbb{R}^p, \mathfrak{B}_i)$, $i = 1, 2$, described respectively, by

$$P_1(\sigma, \sigma^{-1})y = Q_1(\sigma, \sigma^{-1})u$$

and

$$P_2(\sigma, \sigma^{-1})y = Q_2(\sigma, \sigma^{-1})u$$

Let $G_1 = P_1^{-1}Q_1$ and $G_2 = P_2^{-1}Q_2$ denote the transfer functions of these systems. Then $\{\Sigma_1 = \Sigma_2\} \Rightarrow \{G_1 = G_2\}$, but the converse is not always true. In fact:

PROPOSITION 4.9 $\{G_1(s) = G_2(s)\} \Leftrightarrow \{\mathfrak{B}_1^{\text{controllable}} = \mathfrak{B}_2^{\text{controllable}}\}$.

Proof Let $\mathfrak{B}$ be the behaviour of the i/o system described by $P(\sigma, \sigma^{-1})y = Q(\sigma, \sigma^{-1})u$. Then (see section 4.3.1) there exists a $F(s, s^{-1}) \in \mathbb{R}^{p \times p}[s, s^{-1}]$ such that $\det F \neq 0$, $[P \vdots -Q] = F[P' \vdots -Q']$, and $\mathfrak{B}^{\text{controllable}} = \ker [P'(\sigma. \sigma^{-1} \vdots$

$-Q'(\sigma,\sigma^{-1}]$. Clearly $(P')^{-1}Q' = G = P^{-1}Q$. Apply this to (P_1, Q_1) and to (P_2, Q_2). The result follows. ■

There is an obvious relationship between the transfer function and the frequency response of a controllable system (i/o). Let $\mathbb{F}: \mathbb{C} \to \mathscr{G}(m, \mathbb{C}^q)$ be its frequency response, and $G(s) \in \mathbb{R}^{p \times m}(s)$ be its transfer function. Now substitute the indeterminate s by the complex number $0 \neq \lambda \in \mathbb{C}$. If $\lim_{\lambda' \to \lambda} G(\lambda') = \infty$ then we will call λ a *pole* of G. Now $\mathbb{F}(\rho)$ is an m-dimensional subspace of $\mathbb{C}^q$. If e^ρ is not a pole of G then $F(\rho)$ is the graph in $\mathbb{C}^q \cong \mathbb{C}^m \times \mathbb{C}^p$ of the linear map from $\mathbb{C}^m$ to $\mathbb{C}^p$ given by $a \mapsto G(e^\rho)a$.

4.6.5 *On cancelling common factors*

Ever since transfer functions were introduced in control, people have felt uneasy about the question: *is it allowed to cancel common factors in a transfer function*? Our response is a very diplomatic *yes and no*. Indeed, if we look at our definition of the transfer then the answer is *yes—by definition*: in rational functions and matrices of rational functions, common factors can always be cancelled. However, if we look at the behaviour or the frequency response of a system, the answer must be *no*. This ambiguity is not an issue in controllable systems since *in controllable systems there are no common factors*. Let us explain this double talk.

Consider $P_1(s, s^{-1}) \in \mathbb{R}^{g \times q_1}[s, s^{-1}]$ and $P_2(s, s^{-1}) \in \mathbb{R}^{g \times q_2}[s, s^{-1}]$. $L(s, s^{-1}) \in \mathbb{R}^{g \times g}[s, s^{-1}]$ will be called a *common left factor* of P_1 and P_2 if there exist $P'_1(s, s^{-1}) \in \mathbb{R}^{g \times q_1}[s, s^{-1}]$ and $P'_2(s, s^{-1}) \in \mathbb{R}^{g \times q_2}[s, s^{-1}]$ such that $P_1 = LP'_1$ and $P_2 = LP'_2$; we will call the common left factor L *trivial* if it is unimodular; P_1 and P_2 are called *left coprime* if every common left factor is *trivial*. *Common right factors* and *right coprimeness* are defined analogously.

Now consider the system (i/o): $P(\sigma, \sigma^{-1})y = Q(\sigma, \sigma^{-1})u$ with $\det P \neq 0$. Assume that P and Q have L as a common left factor, $P = LP'$, $Q = LQ'$. If this common left factor is trivial, then the input/output systems (i/o): $P(\sigma, \rho^{-1})y = Q(\sigma, \sigma^{-1})u$ and (i/o)$'$: $P'(\sigma, \sigma^{-1})y = Q'(\sigma, \sigma^{-1})u$ have the same behaviour, hence the same frequency response and the same transfer function. If this common left factor is not trivial then (i/o) and (i/o)$'$ have the same transfer function, may or may not have the same frequency response, but they definitely do not have the same behaviour. However, the controllable part of their behaviour is the same. If P and Q are left coprime, then (i/o) and (i/o)$'$ always have the same behaviour as well as the same transfer function and frequency response.

4.6.6 *Factoring the transfer function*

Let $G(s) \in \mathbb{R}^{p \times m}(s)$. Then $G(s)$ admits a *left factorization* as $G(s) = P^{-1}(s, s^{-1})Q(s, s^{-1})$ with $P(s, s^{-1}) \in \mathbb{R}^{p \times p}[s, s^{-1}]$, $Q(s, s^{-1}) \in \mathbb{R}^{p \times m}[s, s^{-1}]$, $\det P \neq 0$. Among these left factorizations there exists one such that P and Q are left coprime. Similarly $G(s)$ admits a *right factorization* as $G(s) = M(s, s^{-1})N^{-1}(s, s^{-1})$ with $M(s, s^{-1}) \in \mathbb{R}^{p \times m}[s, s^{-1}]$, $N(s, s^{-1}) \in \mathbb{R}^{m \times m}[s, s^{-1}]$, $\det N \neq 0$. Among these right factorizations there exists one such that M and N are right coprime.

These factorizations admit natural system theoretic interpretations. A left factorization yields an (AR) representation

$$P(\sigma,\sigma^{-1})y = Q(\sigma,\sigma^{-1})u$$

of an (i/o) system having transfer function G. All systems resulting from such a factorization have the same controllable part. If the factorization is left coprime then this resulting system is controllable. Consequently the left coprime factorization results in the smallest behaviour yielding the desired transfer function: it introduces no responses which are not coded already in the transfer function. Left factorizations which are not coprime introduce *superfluous modes* in the behaviour.

A right factorization yields an (MA) representation

$$u = N(\sigma,\sigma^{-1})a$$
$$y = M(\sigma,\sigma^{-1})a$$

of an (i/o) system having transfer function G. Since it is an (MA) system, this model is always controllable. If the factorization is right coprime, then this system has the property that the latent variable a is observable from

$$\begin{bmatrix} u \\ \cdots \\ y \end{bmatrix}.$$

Hence right factorizations which are not coprime introduce *hidden modes* in the behaviour of the latent variables.

4.7 State models

4.7.1 *Three evolution laws*

In section 1 we have seen that state models will be described by first order difference equations. We will now study finite-dimensional linear time invariant complete state systems. This leads to three model classes. Viewed as equations in the state *and* the signal, the first is of the (AR) type, the second is of the (MA) type, while the third is of the input/output type.

Let us consider a linear evolution law $\Sigma_\partial = (\mathbb{Z}, \mathbb{R}^q, \mathbb{R}^n, \partial)$. Hence ∂ is a linear subspace of $\mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^n$. Expressing ∂ as the kernel of a matrix $[F \vdots G \vdots E]: \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^n \to \mathbb{R}^f$ shows that we are actually looking at an (ARMA) system of the type

$$E\sigma x + Fx + Gw = 0 \qquad \text{(S)}$$

in the variable $x \in \mathbb{R}^n$, the state, and the external variable $w \in \mathbb{R}^q$. (S) relates the state trajectory $x: \mathbb{Z} \to \mathbb{R}^n$ and the signal trajectory $w: \mathbb{Z} \to \mathbb{R}^q$. The model parameters consist of the matrices $E \in \mathbb{R}^{f \times n}$, $F \in \mathbb{R}^{f \times n}$, and $G \in \mathbb{R}^{f \times q}$. The characteristic feature of (S) is that, as a lag relation, it is *first order* in x and *zeroth order* in w. We will first show that (S) defines exactly the class of linear time invariant complete state models. We will call (S) a *state model* (in the category of systems under consideration).

PROPOSITION 4.10 Consider $\Sigma_s=(\mathbb{Z},\mathbb{R}^q,\mathbb{R}^n,\mathfrak{B}_s)$. Then $\{\mathfrak{B}_s\in\mathcal{L}^{q\times n}$ satisfies the axiom of state$\}\Leftrightarrow\{\exists f\in\mathbb{Z}_+$ and matrices $E,F\in\mathbb{R}^{f\times n}$, $G\in\mathbb{R}^{f\times q}$ such that $\mathfrak{B}_s=\ker[G\,\vdots\,E\sigma+F]\}$; $[G\,\vdots\,E\sigma+F]$ is here viewed as a map from $(\mathbb{R}^q\times\mathbb{R}^n)^{\mathbb{Z}}$ into $(\mathbb{R}^f)^{\mathbb{Z}}$. In fact, we can always take $0\leq f\leq q+n$.

Proof ($\Leftarrow$) linearity, shift invariance, and closeness follow from Proposition 4.2. Now the restriction $\operatorname{col}[w,x]\in\ker[G\,\vdots\,E\sigma+F]$ is identical to imposing the evolution law $\partial=\ker[F\,\vdots\,G\,\vdots\,E]$. The state property follows. ($\Rightarrow$) Define $\partial:=\{(a,w,b)\in\mathbb{R}^n\times\mathbb{R}^q\times\mathbb{R}^n\,|\,\exists(w,x)\in\mathfrak{B}_s$ such that $x(0)=a$, $w(0)=w$ and $x(1)=b\}$. Since ∂ is linear there exist matrices E,F,G such that $\partial=\ker[F\,\vdots\,G\,\vdots\,E\,\vdots]$. Now the behaviour induced by this evolution law equals $\ker[G\,\vdots\,\sigma E+F]$. By Theorem 1.1 and the fact that $\mathfrak{B}_s\in\mathcal{L}^{q+n}$ implies that it is complete, we conclude that $\mathfrak{B}_s=\ker[G\,\vdots\,E\sigma+F]$.

We will now indicate why we can always take $f\leq q+n$. Let $X^{\text{eff}}:=\{x\in\mathbb{R}^n\,|\,\exists\,(w,x)\in\mathfrak{B}_s$ such that $x(0)=x\}$. Let ∂ be as defined in ($\Rightarrow$) and let $\partial':=\{(a,w,b)\,|\,a,b\in X^{\text{eff}}$ and $(a,w,b)\in\partial\}$. Denote $n'=\dim X^{\text{eff}}$. Then clearly $\dim\partial'\geq n'$ and hence we need to introduce at most $q+n'$ linear equations on $\mathbb{R}^n\times\mathbb{R}^q\times\mathbb{R}^{n'}$ in order to express ∂' as a kernel. Adding another $n-n'$ equations in order to express that $x\in X^{\text{eff}}$ yields the bound $f\leq q+n'+n-n'=q+n$. ■

Let $\mathcal{S}$ be a linear subspace of $\mathbb{R}^{n_1}\times\mathbb{R}^{n_2}$ having $P_1\mathcal{S}=\mathbb{R}^{n_1}$ where $P_1:(x_1,x_2)\mapsto x_1$. Then $\mathcal{S}$ can always be written as $\mathcal{S}=\{(x_1,x_2)\,|\,x_2\in Lx_1+\mathcal{S}_2\}$ for some matrix $L\in\mathbb{R}^{n_2\times n_1}$ and a subspace $\mathcal{S}_2$ of $\mathbb{R}^{n_2}$. Now apply this to $\partial=\ker[F\,\vdots\,G\,\vdots\,E]$. And assume first that (S) is state trim. It follows that (S) may be written as

$$\begin{bmatrix}\sigma x\\ w\end{bmatrix}=\begin{bmatrix}\tilde{A}\\ \tilde{D}\end{bmatrix}x+\mathcal{L}$$

with $\tilde{A}\in\mathbb{R}^{n\times n}$, $\tilde{C}\in\mathbb{R}^{q\times n}$, and $\mathcal{L}$ a subspace of $\mathbb{R}^{n+q}$. Of course, $\mathcal{L}$ can be written as the image of a matrix, say

$$\mathcal{L}=\operatorname{im}\begin{bmatrix}\tilde{B}\\ \cdots\\ \tilde{D}\end{bmatrix}$$

with $\tilde{B}\in\mathbb{R}^{n\times m}$ and $\tilde{D}\in\mathbb{R}^{q\times m}$. Note that we can always assume that

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

Note that this representation also holds for systems which are not state trim, simply by making sure that $\tilde{A}$ and $\tilde{B}$ map into X^{eff}. This shows that $\mathfrak{B}_s$ can always be expressed as

$$\begin{aligned} \sigma x &= \tilde{A}x + \tilde{B}v \\ w &= \tilde{C}x + \tilde{D}v \end{aligned} \tag{DV}$$

We will call this a model with *driving variables*. In (DV) the variables are $w \in \mathbb{R}^q$, $x \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$; $w: \mathbb{Z} \to \mathbb{R}^q$ is the external signal, $x: \mathbb{Z} \to \mathbb{R}^n$ is the (internal) state trajectory, and $v: \mathbb{Z} \to \mathbb{R}^m$ is the (internal) *driving input*. The matrices $\tilde{A} \in \mathbb{R}^{n \times n}$, $\tilde{B} \in \mathbb{R}^{n \times m}$, $\tilde{C} \in \mathbb{R}^{q \times m}$, and $\tilde{D} \in \mathbb{R}^{q \times m}$ are the parameter matrices of the model. In (DV) the driving input should be regarded as a set of free but latent variables which generate, together with the initial conditions, the state trajectory and the external signal.

Specializing (S) to the case in which the driving variable is a component of the signal leads, finally, to the long-awaited ubiquitous system

$$\begin{aligned} \sigma x &= Ax + Bu \\ y &= Cx + Du \end{aligned} \qquad w = \begin{bmatrix} u \\ \cdots \\ y \end{bmatrix} \tag{i/s/o}$$

This model relates the variables $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, and $y \in \mathbb{R}^p$; $x: \mathbb{Z} \to \mathbb{R}^n$ is the (internal) state trajectory, $u: \mathbb{Z} \to \mathbb{R}^m$ is the input trajectory, and $y: \mathbb{Z} \to \mathbb{R}^p$ is the output trajectory. Together these latter two make up the external signal trajectory

$$w = \begin{bmatrix} u \\ \cdots \\ y \end{bmatrix} : \mathbb{Z} \to \mathbb{R}^q = \mathbb{R}^{p+m}.$$

The matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, and $D \in \mathbb{R}^{p \times m}$ are the parameter matrices of this model, which we call an *input/state/output* system.

4.7.2 *The external behaviour of* (S)

The next proposition shows that (S) gives us one more way of representing elements of $\mathscr{L}^q$.

PROPOSITION 4.11 Let $\mathfrak{B}$ be the external behaviour inducted by (S), i.e., $\mathfrak{B} = P_w \ker [G \vdots E\sigma + F]$ with $P_w: \mathbb{R}^q \times \mathbb{R}^n \to \mathbb{R}^q$ the natural projection. Then $\mathfrak{B} \in \mathscr{L}^q$. Conversely, if $\mathfrak{B} \in \mathscr{L}^q$, then there exists $f \in \mathbb{Z}_+$ and matrices $E, F, G \in \mathbb{R}^{f \times g}$, such that $\mathfrak{B}$ is the external behaviour induced by (S).

Proof The first part is an immediate consequence of section 4.2.3. To show the converse, assume $\mathfrak{B} \in \mathscr{L}^q$. Then by Proposition 4.2 there exists a $R(s) \in \mathbb{R}^{g \times d}[s]$ such that $\mathfrak{B} = \ker R(\sigma)$. Let $R(s) = R_L s^L + R_{L-1} s^{L-1} + \cdots + R_0$. Now consider the following system of equations (suggested by defining as inefficient-state $x := \mathrm{col}[w, \sigma w, \ldots, \sigma^L w]$):

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \sigma x - \begin{bmatrix} I & 0 & 0 & \cdots & 0 \\ 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & I \\ R_0 & R_1 & R_2 & \cdots & R_L \end{bmatrix} x + \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} w = 0$$

and verify that the external behaviour is indeed $\ker R(\sigma)$. ■

It follows immediately that every $\mathfrak{B} \in \mathcal{L}^q$ also admits a representation (DV). We will treat the relation with (i/s/o) later on. In the next two sections we will derive some results which will allow us to check when the state systems (S), (DV), or (i/s/o) are minimal state space realizations of their external behaviour.

4.7.3 *Trimness*

We will first discuss the computation of the effective state space X^{eff} of the model (S). Recall that $X^{\mathrm{eff}} := \{x \in \mathbb{R}^n \mid \exists (w, x)$ satisfying (S) such that $x(0) = x\}$. Consider the following recursive algorithms involving subspaces of the state space $\mathcal{X} = \mathbb{R}^n$.

$$\mathcal{R}_{k+1}^+ = E^{-1}(F\mathcal{R}_k^+ + \mathrm{im}\, G) \qquad \mathcal{R}_0^+ = \mathcal{X}$$
$$\mathcal{R}_{k+1}^- = F^{-1}(E\mathcal{R}_k^- + \mathrm{im}\, G) \qquad \mathcal{R}_0^- = \mathcal{X}$$

(E^{-1} and F^{-1} are inverse images: $\{E^{-1}\mathcal{L} := \{a \mid Ea \in \mathcal{L}\}$). Clearly these linear subspaces satisfy $\mathcal{X} = \mathcal{R}_0^+ \supseteq \mathcal{R}_1^+ \supseteq \cdots \supseteq \mathcal{R}_k^+ \supseteq \cdots$ and $\mathcal{X} = \mathcal{R}_0^- \supseteq \mathcal{R}_1^- \supseteq \cdots \supseteq \mathcal{R}_k^- \supseteq \cdots$ The significance of these spaces will be explained in the proof of the next proposition. From the above inclusion it follows that the limits $\mathcal{R}_\infty^+ := \lim_{t \to \infty} \mathcal{R}_k^+$ and $\mathcal{R}_\infty^- := \lim_{t \to \infty} \mathcal{R}_k^-$ exist and are reached in at most n steps.

PROPOSITION 4.12 $X^{\mathrm{eff}} = \mathcal{R}_\infty^- \cap \mathcal{R}_\infty^+$. Consequently, the system (S) is state trim iff $\mathrm{im}\, E \subseteq \mathrm{im}\, F + \mathrm{im}\, G$ and $\mathrm{im}\, F \subseteq E + \mathrm{im}\, G$

Proof

Clearly	$\mathcal{R}_{k+1}^+$	$= \{a \in \mathbb{R}^n \mid$	$\exists b \in \mathcal{R}_k^+, w \in \mathbb{R}^q$ such that $Ea + Fb + Gw = 0\}$.
Hence	$\mathcal{R}_{k+1}^+$	$= \{x_0 \in \mathbb{R}^n \mid$	$\exists$, for $t = 0, 1, \ldots, k$, $w_t \in \mathbb{R}^q$ and $x_{t+1} \in \mathbb{R}^n$ such that $Ex_{t+1} + Fx_t + Gw_t = 0\}$.
This implies	$\mathcal{R}_\infty^+$	$= \{x_0 \in \mathbb{R}^n \mid$	$\exists$, for $t \geq 0$, $w_t \in \mathbb{R}^q$, $x_{t+1} \in \mathbb{R}^m$, such that $Ex_{t+1} + Fx_t + Gw_t = 0\}$.
Similarly	$\mathcal{R}_{k+1}^-$	$= \{b \in \mathbb{R}^n \mid$	$\exists b \in \mathcal{R}_k^-$, $w \in \mathbb{R}^q$ such that $Ea + Fb + Gw = 0\}$.

Hence $\mathscr{R}^+_{k+1} = \{x_0 \in \mathbb{R}^n \mid \exists$ for $t = 0, 1, \ldots, k$, $w_{-t-1} \in \mathbb{R}^q$ and $x_{-t-1} \in \mathbb{R}^n$ such that $Ex_{-t-1} + Fx_{-t} + Gw_{-t} = 0\}$.

This implies $\mathscr{R}^-_\infty = \{x_0 \in \mathbb{R}^n \mid \exists$, for $t < 0$, $w_t \in \mathbb{R}^q$, $x_t \in \mathbb{R}^n$ such that $Ex_{t+1} + Fx_t + Gw_t = 0\}$.

Consequently $\mathscr{R}^-_\infty \cap \mathscr{R}^+_\infty = \{x_0 \in \mathbb{R}^n \mid \exists (w, x)$ such that $E\sigma w + Fx + Gw = 0$ and $x(0) = x_0\} = X^{\text{eff}}$ ■

It is easy to see when (DV) or (i/s/o) are trim. We state the result formally but without proof.

Proposition 4.13 The system (DV) is state trim iff $\tilde{A}\mathbb{R}^n + \operatorname{im}\tilde{B} = \mathbb{R}^n$. The system (i/s/o) is state trim iff $A\mathbb{R}^n + \operatorname{im} B = \mathbb{R}^n$.

4.7.4 *Past- and future-induced*

In order to discuss the question when the state in (S), (DV), or (i/s/o) is past or future induced, it will be convenient to introduce the notion of an observable pair of matrices. A pair of matrices (M, N), $M \in \mathbb{R}^{n \times n}$, $N \in \mathbb{R}^{p \times n}$ is said to be an *observable pair* if $\{x(t+1) = Mx(t);\ Nx(t) = 0$ for $t \in \mathbb{Z}_+\} \Rightarrow \{x(0) = 0\}$. This is equivalent to requiring that the largest M-invariant subspace contained in $\ker N$ is zero. Equivalently, as a matrix test, that $\operatorname{rank}\operatorname{col}[N, NM, \ldots, NM^{n-1}] = 0$. Stated in yet another way:

$$\operatorname{rank}_{\mathbb{C}}\begin{bmatrix} I\lambda - M \\ \cdots\cdots \\ N \end{bmatrix} = n \qquad \text{for all } \lambda \in \mathbb{C}.$$

Specializing the notion of past and future induced (see section 2) to linear systems shows that the state in a linear state space system will be past induced iff $\{(w, x) \in \mathfrak{B}_s,\ w(t) = 0$ for $t < 0\} \Rightarrow \{x(0) = 0\}$ and future induced iff $\{(w, x) \in \mathfrak{B}_s,\ w(t) = 0$ for $t \geq 0\} \Rightarrow \{x(0) = 0\}$. In order to verify if these conditions are satisfied, consider the following recursions:

$$\mathscr{V}^+_{k+1} = E^{-1}F\mathscr{V}^-_k \qquad \mathscr{V}^+_0 = \mathscr{X}$$
$$\mathscr{V}^-_{k+1} = F^{-1}E\mathscr{V}^-_k \qquad \mathscr{V}^-_0 = \mathscr{X}$$

Clearly these linear subspaces satisfy $\mathscr{X} = \mathscr{V}^+_0 \geq \mathscr{V}^+_1 \geq \cdots \geq \mathscr{V}^+_k \geq \cdots$ and $\mathscr{X} = \mathscr{V}^-_0 \geq \mathscr{V}^-_1 \geq \cdots \geq \mathscr{V}^-_k \geq \cdots$ The significance of these spaces will be explained in the proof of the next proposition. From the above inclusions it follows that the limits $\mathscr{V}^+_\infty := \lim_{k\to\infty} \mathscr{V}^+_k$ and $\mathscr{V}^-_\infty := \lim_{k\to\infty} \mathscr{V}^-_k$ exist and are reached in at most n steps.

Proposition 4.14 Assume that (S) is state trim. Then the state in (S) is past induced iff $\mathscr{V}^-_\infty = 0$ and future induced iff $\mathscr{V}^+_\infty = 0$.

Proof

Clearly $\mathscr{V}_{k+1}^{+} = \{a \in \mathbb{R}^n \mid \exists b \in \mathscr{V}_k^{+}$ such that $Ea + Fb = 0\}$

Hence $\mathscr{V}_{k+1}^{+} = \{x_0 \in \mathbb{R}^n \mid \exists$, for $t = 0, 1, \ldots, k$, $x_{t+1} \in \mathbb{R}^n$ such that $Ex_{t+1} + Fx_t = 0\}$

This implies $\mathscr{V}_{\infty}^{+} = \{x_0 \in \mathbb{R}^n \mid \exists$, for $t \geq 0$, x_{t+1} such that $Ex_{t+1} + Fx_t = 0\}$

The result for $\mathscr{V}_{\infty}^{+}$ follows, and for $\mathscr{V}_{\infty}^{-}$ it is proven analogously. ■

Recall that a state behaviour $\mathfrak{B}_s$ is called *deterministic* if the value of the state and the signal at time t uniquely determine the state at time $t+1$. For linear systems this requires $\{(w, x) \in \mathfrak{B}_s, w(0) = 0, x(0) = 0\} \Rightarrow \{x(1) = 0\}$. Clearly if (S) is state trim, then it will be deterministic iff $\ker E = 0$. The system (DV) is deterministic iff $\ker D \subseteq \ker B$. The system (i/s/o) is always deterministic.

It is easy to see that for (S) there holds: {the state is past and future induced} $\Leftrightarrow$ {the state is future induced and deterministic}.

4.7.5 *Minimality and reduction*

We will now discuss the problem of deducing a minimal state space models (S) representing an external behaviour $\mathfrak{B} \in \mathscr{L}^q$. However, before turning to this problem, it is instructive to examine first how much can be deduced in the linear case from the abstract constructions in section 2.

Let $\Sigma = (T, W; \mathfrak{B})$ be the external behaviour of the linear state system $\Sigma_s = (T, W, X, \mathfrak{B}_s)$. Then, as we have seen, Σ_s will be a minimal state representation if and only if Σ_s is state trim and if the state is both past and future induced.

If Σ_s is not minimal then a minimal state representation can be constructed as follows. Compute $X^{\text{eff}} := \{x \in X \mid \exists (w, x) \in \mathfrak{B}_s$ such that $x(0) = x\}$, $\mathscr{V}^{+} := \{x \in X \mid \exists (w, x) \in \mathfrak{B}_s$ such that $x(0) = x$ and $w(t) = 0$ for $t \geq 0\}$, and $\mathscr{V}^{-} := \{x \in X \mid \exists (w, x) \in \mathfrak{B}_s$ such that $x(0) = x$ and $w(t) = 0$ for $t < 0\}$. Clearly $\mathscr{V}^{-} \subset X^{\text{eff}}$, $\mathscr{V}^{+} \subset X^{\text{eff}}$. Now define $X^{\text{red}} := X^{\text{eff}} (\text{mod}\,(\mathscr{V}^{-} + \mathscr{V}^{+}))$, $\mathfrak{B}_s^{\text{red}} := \{(w, \tilde{x}) : T \to W \times X^{\text{red}} \mid \exists x$ such that $(w, x) \in \mathfrak{B}_s$ and $\tilde{x}(t) = x(t) (\text{mod}\,(\mathscr{V}^{-} + \mathscr{V}^{+}))$ for all $t \in T\}$. It can be shown that the reduced system $\Sigma^{\text{red}} := (T, W, X^{\text{red}}, \mathfrak{B}_s^{\text{red}})$ is a linear state space representation of Σ, indeed a minimal one.

All minimal linear state representations can be constructed from one as follows. Let $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ be a minimal state space representation of its own external behaviour Σ. Now take any other linear space X', isomorphic to X in the sense that there exist a linear bijection $X \underset{S^{-1}}{\overset{S}{\rightleftharpoons}} X'$. Then $\Sigma_s' = (T, W, X', \mathfrak{B}_s')$, with $\mathfrak{B}_s' = \{(w, x') : T \to W \times X' \mid (w, S^{-1} \circ x') \in \mathfrak{B}_s\}$, is also a minimal linear state representation of Σ and all minimal linear state representations are obtained this way. This fact, together with the above reduction procedure, shows that every linear state representation $\Sigma_s' = (T, W, X, \mathfrak{B}_s)$ is related to a minimal linear state representation $\Sigma_s^{\min} = (T, W, X^{\min}, \mathfrak{B}_s^{\min})$ of the same external behaviour as follows. There exist a surjective linear map S from a linear subspace X' of X to

$X^{\min}$ such that $\{(w, x') \in \mathfrak{B}'_s\} \Leftrightarrow \{\exists x$ such that $(w, x) \in \mathfrak{B}_s$, $x(t) \in X'$ for all $t \in T$, and $x' = S \circ x\}$. If there exists a subspace X_1 such that $X = X^{\text{eff}} \oplus X_1$ then S may be taken to be defined on all of X. If Σ_s is also minimal then S must be a bijection from X into X'.

This allows the following important conclusion. If a linear system $\Sigma = (T, W, \mathfrak{B})$ has a linear state representation $\Sigma_s = (T, W, X, \mathfrak{B}_s)$ with $\dim X < \infty$, then it is a minimal state representation iff $\dim X$ is as small as possible. Hence in this case *we can interpret minimality* (which in Section 2 was defined in a purely set theoretic sense) *simply as 'having a state space of minimal possible dimension'*.

When a state space system is a minimal state representation of its own external behaviour, we will simply call it *minimal*. Let us now examine when (S), (DV), or (i/s/o) are minimal. The results obtained in sections 4.7.3 and 4.7.4 allow us to be very concrete about this.

THEOREM 4.2

(1) (S) *is minimal* $\Leftrightarrow$

$$\left.\begin{array}{ll}\text{(i)} & \operatorname{im} E \subseteq \operatorname{im} F + \operatorname{im} G \\ \text{(ii)} & \operatorname{im} F \subseteq \operatorname{im} E + \operatorname{im} G\end{array}\right\} \qquad (\textit{trim})$$

(iii) $\ker E = 0$ (*determinism*)

(iv) $\mathscr{V}^+_\infty = 0$ (*future induced*)

(2) (DV) *is minimal* $\Leftrightarrow$

(i) $\tilde{A}\mathbb{R}^n + \operatorname{im} \tilde{B} = \mathbb{R}^n$ (*trim*)

(ii) $\ker \tilde{D} \subseteq \ker \tilde{B}$ (*determinism*)

(iii) $(\tilde{A} - \tilde{B}^\# \tilde{C}, \tilde{C}^\#)$ *is an observable pair* (*future induced*)

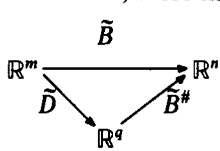
Here $\tilde{B}^\#$ *is any matrix such that* $\tilde{B}^\# \tilde{D} = \tilde{B}$ and $\tilde{C}^\# = P\tilde{C}$, *where* P *denotes the natural projection* $P: \mathbb{R}^q \to \mathbb{R}^q (\operatorname{mod} \operatorname{im} \tilde{D})$.

(3) (i/s/o) *is minimal* $\Leftrightarrow$

(i) $A\mathbb{R}^n + \operatorname{im} B = \mathbb{R}^n$ (*trim*)

(ii) (A, C) *is an observable pair* (*future induced*)

Proof From section 2 we know that {minimal} $\Leftrightarrow$ {trim, past and future induced}. We know from section 4.7.5 that this is equivalent to {trim deterministic, and future induced}. The results of sections 4.7.4 and 4.7.5 show that we need only prove that $\{(\tilde{A} - \tilde{B}^\# \tilde{C}, \tilde{C}^\#)$ is an observable pair$\} \Leftrightarrow \{$the state in (DV) is future induced$\}$. Note that, since $\ker \tilde{D} \subseteq \ker \tilde{B}$, there indeed exists a $\tilde{B}^\#$ such that the diagram



commutes. Now assume that (w, x, v) satisfies (DV) this implies $\tilde{B}^{\#}w = \tilde{B}^{\#}\tilde{C}x + \tilde{B}v$, hence $\sigma x = (\tilde{A} - \tilde{B}^{\#}\tilde{C})x + \tilde{B}^{\#}w$; $w = \tilde{C}x + \tilde{D}v$. Hence (w, x) belongs to the behaviour of (DV) and $w(t) = 0$ for $t \geq 0$ iff, for some $v \in (\mathbb{R}^m)^{\mathbb{Z}_+}$, there holds for all $t \geq 0: x(t+1) = (\tilde{A} - \tilde{B}^{\#}\tilde{C})x(t)$; $0 = \tilde{C}x(t) + \tilde{D}v(t)$ Equivalently iff for $t \geq 0$ there holds $x(t+1) = (\tilde{A} - \tilde{B}^{\#}\tilde{C})x(t)$, $\tilde{C}^{\#}x(t) = 0$. This, however, implies $x(0) = 0$ iff $(\tilde{A}^{\#} - \tilde{B}^{\#}\tilde{C}, \tilde{C}^{\#})$ is observable. We conclude that this observability condition is indeed equivalent to future induced. ■

4.8 Input/state/output systems

4.8.1 *Just as the state is always there, so is the input!*

THEOREM 4.3 *Consider $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ with $\mathfrak{B} \in \mathscr{L}^q$. Then there exists a componentwise partition of $\mathbb{R}^q$ into $\mathbb{R}^q = \mathbb{R}^m \times \mathbb{R}^p$, an integer $n \in \mathbb{Z}_+$, and matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, and $D \in \mathbb{R}^{p \times m}$ such that the input/state/output system*

$$\begin{aligned} \sigma x &= Ax + Bu \\ w &= Cx + Du \end{aligned} \qquad w = \begin{bmatrix} u \\ \cdots \\ y \end{bmatrix} \tag{i/s/o}$$

has $\mathfrak{B}$ as its external behaviour.

Proof $\mathfrak{B} \in \mathscr{L}^q$ implies that it is an (AR) system, which implies, by Proposition 4.9, that it has a representation (S) or, equivalently, (DV). Consequently, it has a minimal representation (DV) and hence a deterministic one, that is, one with $\ker \tilde{D} \subseteq \ker \tilde{B}$. On the other hand, we can always assume, as far as the behaviour $\mathfrak{B}_s$ of (DV) is concerned, that

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

In conclusion, there exists a representation (DV) with $\ker \tilde{D} = 0$. Now reorder the components of w, if need be, and partition the vector w such that the equations for (DV) look like

$$\sigma x = \tilde{A}x + \tilde{B}v, \quad w_1 = \tilde{C}_1 x + \tilde{D}_1 v, \quad w_2 = \tilde{C}_2 x + \tilde{D}_2 v$$

with $\tilde{D}_1$ invertible. This yields

$$\sigma x = (\tilde{A} - \tilde{B}\tilde{D}_1^{-1}\tilde{C}_1)x + \tilde{B}\tilde{D}_1^{-1}w_1, \quad w_2 = (\tilde{C}_2 - \tilde{D}_2\tilde{C}_1^{-1}\tilde{C}_1)x + \tilde{D}_2\tilde{D}_1^{-1}w_1$$

Now observe that the equation $w_1 = \tilde{C}_1 x + \tilde{D}_1 v$ implies that w_1 is free. Next, call $w_1 = u$, $w_2 = y$, and conclude that the input/state/output system (i/s/o), with $A = \tilde{A} - \tilde{B}\tilde{D}_1^{-1}\tilde{C}_1$; $B = \tilde{B}\tilde{D}_1^{-1}$, $C = \tilde{C}_2 - \tilde{D}_2\tilde{D}_1^{-1}\tilde{C}_1$, and $D = \tilde{D}_2\tilde{D}_1^{-1}$, has the same behaviour $\mathfrak{B}_s$ as (DV). ■

The system i/s/o is the starting point of many studies in linear system theory. What we have shown here is that every linear time invariant complete behaviour may be written in this form. *Just as the state is always there*, perhaps and usually implicitly, *so is the input and so is the output* perhaps and sometimes implicitly: *all we need to do is choose appropriately the components of the signal vector* w!

We view the above theorem as important for dynamic simulations. In constructing a model (for example of an electrical or a mechanical system) by tearing and zooming in on subsystems, one will obtain a difference or a differential equation with latent variables consisting of difference or differential equations perhaps of high order, coupled with static constraints (these are usually called algebraic constraints). When attempting to perform a dynamic simulation we need to know which signals need to be selected on the basis of considerations outside of the model, and which initial conditions should be provided. If we assume that the system is linear time invariant and complete and if, for simplicity, we consider the difference equation case, then we have given a procedure of how to approach this. The model with latent variables will be an (ARMA) model. Eliminating the latent variables will yield an (AR) model, which by Theorem 4.3 we can describe in (i/s/o) form. This model is eminently suited for simulation: $u=w_1$ is to be chosen, $x(0)$ should be provided, $y=w_2$ will be computed. Note that it is important to use minimal systems. Otherwise, the initial states outside X^{eff} will introduce *ghost solutions*: elements not present in $\mathfrak{B}^{0+}$. Further, if $\mathcal{V}\neq 0$, there are initial states yielding state trajectories such that $w(t)=0$ for $t\geq 0$. We call these *phantom trajectories*. The nature and the numerical behaviour of these trajectories are not inherent in $\mathfrak{B}^{0+}$. Further, the initial state may be deduced uniquely from $w(t)$, $t<0$, iff $\mathcal{V}^-=0$. This will yield a procedure for choosing $x(0)$. Finally, if $\mathcal{V}^++\mathcal{V}^-\neq 0$, then the unnecessary high dimension of the state space may increase the computational complexity of the simulation.

4.8.2 *Controllability*

A pair of matrices (M,N), $M\in\mathbb{R}^{n\times n}$, $N\in\mathbb{R}^{n\times m}$ is said to be a *controllable pair* if for all x', $x''\in\mathbb{R}^n$ there exists a $t\in\mathbb{Z}_+$ and $u_0,u_1,\ldots,u_{t-1}\in\mathbb{R}^n$ such that $x_{k+1}=Mx_k+Nu_k$ and $x_0=x'$, yield $x_t=x''$. This is equivalent to requiring that the smallest M-invariant subspace containing $\operatorname{im} N$ is $\mathbb{R}^n$. Equivalently, as a matrix test, that $\operatorname{rank}[N,MN,\ldots,M^{n-1}N]=n$. Stated in yet another way: $\operatorname{rank}_{\mathbb{C}}[I\lambda-M\,\vdots\,N]=n$ for all $\lambda\in\mathbb{C}$.

The following result is an immediate consequence of Propositions 1.3 and 2.7.

PROPOSITION 4.15 The external behaviour of (DV) is controllable if $(\tilde{A},\tilde{B})$ is a controllable pair. The external behaviour of (i/s/o) is controllable if (A,B) is a controllable pair. If (DV) or (i/s/o) is controllable if (A,B) is a controllable pair. If (DV) or (i/s/o) are minimal, then these ifs become iffs.

Note, as stated in Theorem 4.2 that minimality of (i/s/o) does not require controllability of (A,B). It is worth emphasizing this because it is contrary to the *dogma*: {minimality}$\Leftrightarrow${controllability and observability} which has been obtained for the classical input/output map type systems [9]. Since our framework incorporates autonomous systems very comfortably, the lack of controllability should come as no surprise: autonomous systems are unaffected by the external world and can hence not have free inputs. However, for controllable systems we can indeed prove:

PROPOSITION 4.16 If the external behaviour of (i/s/o) is controllable, then (i/s/o) is minimal iff (A, B) is controllable and (A, C) is observable.

Proof (only if): follows from Theorem 4.2 and the above proposition. (if): Observe that $\{(A,B)$ controllable$\}\Rightarrow\{\text{im}\,[I\lambda - A \,\vdots\, B] = \mathbb{R}^n$ for all $\lambda \in \mathbb{C}\} \Rightarrow \{\text{lim}\,[A \,\vdots\, B] = \mathbb{R}^n\} \Leftrightarrow \{\text{trim}\}$. Now apply Theorem 4.2 ■

4.8.3 *Integer invariants*

We have seen that we can look upon behaviours $\mathfrak{B} \in \mathscr{L}^q$ as being parametrized in many different ways: as (AR) systems by polynomial matrices $R(s, s^{-1})$; as (i/o) systems by pairs of polynomial matrices $(P(s, s^{-1}), Q(s, s^{-1}))$; as state systems (S) by triples of matrices (E, F, G); as systems with driving variables (DV) or as (i/s/o) systems by quadruples of matrices $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ or (A, B, C, D), respectively. Such systems will hence be characterized on the one hand by some integers: the degrees of the polynomial matrices and the size of the matrices, and on the other hand by real numbers: the coefficients of the matrix polynomials or the entries of the matrices. We will refer to these integers, somewhat informally, as *integer invariants*. They determine the structure of the system whereas the coefficients or the entries determine the parameters. The study of these integer invariants is a rich and appealing subject. We will limit our attention to a few highlights which we will state without proof.

Consider $(\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ with $\mathfrak{B} \in \mathscr{L}^q$. Recall that there will exist a $g \in \mathbb{Z}_+$ and an $R(s, s^{-1}) \in \mathbb{R}^{g \times q}[s, s^{-1}]$ such that $\mathfrak{B} = \ker R(\sigma, \sigma^{-1})$. Let $g_{\min}$ be the minimal such g. Further, there exists a $r \in \mathbb{Z}_+$ and a matrix $M \in \mathbb{R}^{r \times q}$ such that $M \circ \mathfrak{B} = (\mathbb{R}^r)^{\mathbb{Z}}$ (i.e. M filters free variables out of w). Let $r^{\max}$ be the maximal such r. Finally, let $f_{\min}$ be the minimal f such that a model (S) exists having external behaviour $\mathfrak{B}$, and let $m_{\min}$ be the minimal m such that a model (DV) exists having external behaviour $\mathfrak{B}$. There holds:

$$
\begin{aligned}
g_{\min} &= q - r^{\max} \\
&= q - m_{\min} \\
&= q - \text{the number of input variables in any representation of } \mathfrak{B} \text{ in the form (i/o) or (i/s/o)} \\
&= \text{the number of output variables in any representation of } \mathfrak{B} \text{ in the form (i/0) or (i/s/o)}
\end{aligned}
$$

Define the *degree*, $\partial(M)$, of $M(s, s^{-1}) \in \mathbb{R}^{n_1 \times n_2}[s, s^{-1}]$, $M(s, s^{-1}) = M_L s^L + M_{L-1} s^{L-1} + \cdots + M_{l+1} s^{l+1} + M_l s^l$, $M_L \neq 0$, $M_l \neq 0$, to be $L - l$. The *McMillan degree* of the full row rank polynomial matrix $M(s,s^{-1})\in\mathbb{R}^{n_1\times n_2}[s,s^{-1}]$ is defined to be the degree of the vector formed by all its minors. Let $R(s,s^{-1})\in\mathbb{R}^{d\times q}[s,s^{-1}]$ be such that $\mathfrak{B}=\ker R(\sigma,\sigma^{-1})$. Now write out the g lag equations in $R(\sigma,\sigma^{-1})w=0$ individually row by row, yielding

$$\begin{gathered} r_1(\sigma,\sigma^{-1})w=0 \\ r_2(\sigma,\sigma^{-1})w=0 \\ \vdots \\ r_g(\sigma,\sigma^{-1})w=0 \end{gathered}$$

Define $\partial(R)=:$ the *lag* of $R(\sigma,\sigma^{-1})w=0$ and $\sum_{k=1}^{g}\partial(r_k)=:$ the *total lag* of $R(\sigma,\sigma^{-1})w=0$. Define $L_{\min}:=\min\partial(R)$ and $L_{\min}^{\text{tot}}:=\min\sum_{k=1}^{g}\partial(r_k)$ where these minima are taken over all R's such that $\mathfrak{B}=\ker R(\sigma,\sigma^{-1})$. Also define $n_{\min}$ to the minimal $n\in\mathbb{Z}_+$ for which there exists a state space representation (S), (DV) or (i/s/o) of $\mathfrak{B}$. There holds:

The McMillan degree of any $R(s,s^{-1})\in\mathbb{R}^{g_{\min}\times q}[s,s^{-1}]$ such that $\mathfrak{B}=\ker R(\sigma,\sigma^{-1})$

$$\begin{aligned} &= L_{\min}^{\text{tot}} \\ &= n_{\min} \end{aligned}$$

Furthermore,

$$f_{\min}=q-m_{\min}+n_{\min}$$

Recall that in section 1.4.2 we have defined the *memory span* $\Delta_{\min}$ of $\mathfrak{B}$ to be the minimal $\Delta\in\mathbb{Z}_+$ such that $\{w_1,w_2\in\mathfrak{B}, w_1|_{[0,\Delta)}=w_2|_{[0,\Delta)}\}\Rightarrow\{w_1\wedge_0 w_2\in\mathfrak{B}\}$. Finally define the *observability index* ν of an observable pair (M,N), $M\in\mathbb{R}^{n\times n}$, $N\in\mathbb{R}^{p\times n}$, to be the smallest integer k such that $\operatorname{rank}\operatorname{col}[M,NM,\ldots,NM^{k-1}]=n$. Then

$\Delta_{\min}=L_{\min}$

$=$ the observability index of $(\tilde{A}-\tilde{B}^{\#}\tilde{C},\tilde{C}^{\#})$ or (A,C) is any minimal state space representation (DV) or (i/s/o) of $\mathfrak{B}$.

4.8.4 *The feedback group*

We will call models (DV) with $m=m_{\min}$ and $n=n_{\min}$ *minimal state/minimal driving input models.* (DV) defines such a model iff (see Theorem 4.2):

(i) $\tilde{A}\mathbb{R}^n+\operatorname{im}\tilde{B}=\mathbb{R}^n$
(ii) $\ker\tilde{D}=0$
(iii) $(\tilde{A}-\tilde{B}^{\#}\tilde{C},\tilde{C}^{\#})$ is an observable pair.

Note that if (DV) is a minimal state/minimal driving input representation then (x,u) in (DV) is observable from w implying that there exist $F_x(s,s^{-1})\in\mathbb{R}^{n\times q}[s,s^{-1}]$ and $F_v(s,s^{-1})\in\mathbb{R}^{m\times q}[s,s^{-1}]$ such that $\{(w,x,v)\text{ satisfies (DV)}\}\Leftrightarrow\{w\in\mathfrak{B}, x=F_x(\sigma)w,\text{ and } v=F_v(\sigma)w\}$. By the same token we can take F_x and F_v to be polynomials in s or in s^{-1}, with, moreover, $F_x(s^{-1})$ such that it has no constant term.

It is possible to obtain all minimal state/minimal driving input models for a given external behaviour, starting from one. Consider the transformation group $G = \mathcal{G}l(n) \times \mathbb{R}^{n \times m} \times \mathcal{G}l(m)$ acting on $\mathbb{R}^{(n+q)\times(n+m)}$ as follows:

$$(A, B, C, D) \underset{\substack{S \in \mathcal{G}l(n) \\ F \in \mathbb{R}^{n \times m} \\ R \in \mathcal{G}l(m)}}{\overset{(S,F,R)}{\longmapsto}} (S(A+BF)S^{-1}, SBR, (C+DF)S^{-1}, DR)$$

It is easy to see that this transformation group leaves the external behaviour of the system (DV) invariant; it leaves also the minimal state/minimal driving input elements invariant. The orbit under G of one minimal state/minimal driving input element generates in fact exactly all the minimal state/minimal driving elements with the same external behaviour. The transformation group G has been the object of much study in the mathematical system theory literature under the name of the *feedback group*. The above shows that this transformation group plays an extremely natural role in the classification of state representations (DV): it corresponds to leaving the external behaviour invariant.

4.9 Wrap-up

We will now summarize without further proofs or comments the main results obtained in this section in a series of three theorems. The first one treats the general situation, the second the reachable case, while the third and final theorem treats autonomous systems.

4.9.1 *Linear time invariant systems*

THEOREM 4.4 *Let $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ be a dynamical system. Then the following conditions are equivalent:*

(1) *Σ is linear, time invariant, and complete;*
(2) *$\mathfrak{B} \in \mathscr{L}^q$, i.e., $\mathfrak{B}$ is linear shift invariant closed subspace of $(\mathbb{R}^q)^{\mathbb{Z}}$;*
(3) *$\exists g \in \mathbb{Z}_+$ and a polynomial matrix $R(s, s^{-1}) \in \mathbb{R}^{g \times q}[s, s^{-1}]$ such that $\mathfrak{B} = \ker R(\sigma^{-1})$, i.e., $\mathfrak{B}$ admits an (AR) representation;*
(4) *$\exists f, d \in \mathbb{Z}_+$ and polynomial matrices $R'(s, s^{-1}) \in \mathbb{R}^{f \times q}[s, s^{-1}]$, $R''(s, s^{-1} \in \mathbb{R}^{f \times d}$ $[s, s^{-1}]$ such that $\mathfrak{B} = (R'(\sigma, \sigma^{-1}))^{-1} \operatorname{im} R''(\sigma, \sigma^{-1})$, i.e., $\mathfrak{B}$ admits an $(ARMA)$ representation;*
(5) *$\exists p, m \in \mathbb{Z}_+$, $p + m = q$, a permutation matrix $\pi \in \mathbb{R}^{q \times q}$, and polynomial matrices $P(s, s^{-1}) \in \mathbb{R}^{p \times p}[s, s^{-1}]$, $Q(s, s^{-1}) \in \mathbb{R}^{p \times m}[s, s^{-1}]$ with $P^{-1}(s, s^{-1})Q(s, s^{-1}) \in \mathbb{R}_+^{p \times m}(s)$ such that $\mathfrak{B} = \pi \ker [P(\sigma, \sigma^{-1}) \vdots -Q(\sigma, \sigma^{-1})]$, i.e., $\mathfrak{B}$ admits a componentwise (i/o) representation which is non-anticipating;*
(6) *$\exists n \in \mathbb{Z}_+$ such that $\mathfrak{B}$ is the external behaviour of a state space system $\Sigma_s = (\mathbb{Z}, \mathbb{R}^q, \mathbb{R}^n, \mathfrak{B}_s)$ with $\mathfrak{B}_s \in \mathscr{L}^{q \times n}$ i.e., $\mathfrak{B}$ admits a finite-dimensional linear time-invariant state realization;*
(7) *$\exists f, n \in \mathbb{Z}_+$ and matrices $E, F \in \mathbb{R}^{f \times n}$, $G \in \mathbb{R}^{f \times g}$ such that $\mathfrak{B} = P_w \ker [G \vdots \sigma E + F]$, i.e., $\mathfrak{B}$ admits a representation (S);*
(8) *$\exists n, m \in \mathbb{Z}_+$ and matrices $\tilde{A} \in \mathbb{R}^{n \times m}$, $\tilde{B} \in \mathbb{R}^{n \times m}$, $\tilde{C} \in \mathbb{R}^{q \times m}$, $\tilde{D} \in \mathbb{R}^{q \times m}$ such*

that $\mathfrak{B} = \{w \mid \exists x, y$ *such that* $\sigma x = \tilde{A}x + \tilde{B}v,\ w = \tilde{C}x + \tilde{D}v\}$ i.e., $\mathfrak{B}$ *admits a representation* (*DV*);

(9) $\exists m, n, \in \mathbb{Z}_+$, $m + p = q$, *a permutation matrix* $\pi \in \mathbb{R}^{q \times q}$, *and matrices* $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, $D \in \mathbb{R}^{p \times m}$ *such that* $\pi^{-1}\mathfrak{B} = \{(u, y) \mid \exists x$ *such that* $\sigma x = Ax + Bu,\ y = Cx + Du\}$, *i.e.,* $\mathfrak{B}$ *admits a componentwise* (*i/s/o*) *representation.*

4.9.2 *Controlable systems*

THEOREM 4.5 *Let* $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ *be a dynamical system. Then the following conditions are equivalent:*

(1) Σ *is linear, time invariant, complete, and controllable;*

(2) $\mathfrak{B} \in \mathscr{L}^q$ *and* $\mathfrak{B} = \mathfrak{B}^{\text{compact/closure}}$;

(3) $\mathfrak{B}$ *admits an* (*AR*) *representation with* $\ker_{\mathbb{C}} R(\lambda, \lambda^{-1})$ *of constant dimension for* $0 \neq \lambda \in \mathbb{C}$;

(4) $\exists d \in \mathbb{Z}_+$ *and a polynomial matrix* $M(s, s^{-1}) \in \mathbb{R}^{q \times d}[s, s^{-1}]$ *such that* $\mathfrak{B} = \operatorname{im} M(\sigma, \sigma^{-1})$, *i.e.,* $\mathfrak{B}$ *admits an* (*MA*) *rrepresentation;*

(5) $\exists p, m \in \mathbb{Z}_+$, $p + m = q$, *a permutation matrix* $\pi \in \mathbb{R}^{q \times q}$, *and left coprime polynomial matrices* $P(s, s^{-1}) \in \mathbb{R}^{p \times p}[s, s^{-1}]$, $Q(s, s^{-1}) \in \mathbb{R}^{p \times m}[s, s^{-1}]$ *with* $P^{-1}(s, s^{-1})Q(s, s^{-1}) \in \mathbb{R}_+^{p \times m}(s)$, *such that* $\mathfrak{B} = \pi[P(\sigma, \sigma^{-1}) \vdots -Q(\sigma, \sigma^{-1})]$;

(6) $\exists p, m \in \mathbb{Z}_+$, $p + m = q$, *a permutation matrix* $\pi \in \mathbb{R}^{q \times q}$, *and right coprime polynomial matrices* $M(s, s^{-1}) \in \mathbb{R}^{p \times m}[s, s^{-1}]$, $N(s, s^{-1}) \in \mathbb{R}^{m \times m}[s, s^{-1}]$ *with* $M(s, s^{-1})N^{-1}(s, s^{-1}) \in \mathbb{R}_+^{p \times m}(s)$, *such that*

$$\mathfrak{B} = \pi \operatorname{im} \begin{bmatrix} N(\sigma, \sigma^{-1}) \\ \cdots\cdots \\ M(\sigma, \sigma^{-1}) \end{bmatrix}$$

(7) $\mathfrak{B}$ *admits a representation* (*i/s/o*) *with* (*A, B*) *a controllable pair.*

4.9.3 *Autonomous systems*

THEOREM 4.6. *Let* $\Sigma = (\mathbb{Z}, \mathbb{R}^q, \mathfrak{B})$ *be a dynamical system. Then the following conditions are equivalent:*

(1) Σ *is a linear, time invariant, complete, and autonomous;*

(2) $\mathfrak{B}$ *is a finite-dimensional shift invariant subspace of* $(\mathbb{R}^q)^{\mathbb{Z}}$;

(3) $\mathfrak{B}$ *admits an* (*AR*) *representation with*

$$R(s, s^{-1}) \in \mathbb{R}^{q \times q}[s, s^{-1}] \quad \textit{and} \quad \det_{\mathbb{R}(s)} R(s, s^{-1}) \neq 0.$$

(4) $\mathfrak{B}$ *admits a representation* (*DV*) *with* $m = 0$ *i.e., a representation* $\sigma w = Ax$, $w = Cx$.

4.9.4 *Symbolic calculus*

The results obtained in this section can easily be generalized, *mutatis mutandis*, to the case that the time axis is $\mathbb{Z}_+$. *Mutatis mutandis* here means: interpret time invariance and shift invariance as $\sigma\mathfrak{B} \subseteq \mathfrak{B}$ and work throughout with $\mathbb{R}[s]$ instead of with $\mathbb{R}[s, s^{-1}]$.

We have concentrated in our exposition on the case $T = \mathbb{Z}$. In [14] we have attempted to treat the cases $T = \mathbb{Z}$ and $T = \mathbb{Z}_+$ in parallel. The present exposition has pedagogical advantages, but it has the important disadvantage that we do not quite cover the continuous time case, since that case is basically identical to the case $T = \mathbb{Z}_+$. In order to translate the results from $T = \mathbb{Z}_+$ to $T = \mathbb{R}$ or $\mathbb{R}_+$, simply interpret the shift σ as the differential operator d/dt. We can view this interpretation as *symbolic calculus*. In particular, Proposition 4.1c will imply that if a system is governed by a set of linear differential equations involving latent variables:

$$R_1\left(\frac{d}{dt}\right)\boldsymbol{w} = R_2\left(\frac{d}{dt}\right)\boldsymbol{a},$$

then there will exist a polynomial matrix $R(s)$ such that w is governed by a set of linear differential equations

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

As far as the smoothness is concerned in this result, we can either assume that both $\boldsymbol{w}$ and $\boldsymbol{a}$ are C^∞, alternatively that $\boldsymbol{w}$ and $\boldsymbol{a}$ are both distributions, or finally that $\boldsymbol{w} \in \mathscr{L}^{\text{loc}}(\mathbb{R}; \mathbb{R}^q)$, that $\boldsymbol{a}$ is a distribution, and that the differential equations are satisfied in the sense of distributions. The C^∞ case is easy to work with mathematically but has the conceptual disadvantage that differential equations which are first order in the auxiliary variable x and zeroth order in w will then not satisfy the axiom of state. Considering all equations in the sense of distributions and assuming the behaviour $\mathfrak{B}$ to be a family of distributions is by and large the most convenient approach, however.

As we have already said, all the results remain essentially valid, after suitable modifications, in the case $T = \mathbb{Z}_+$, $\mathbb{R}$ or $\mathbb{R}_+$. One difference occurs in the condition for trimness: in particular, (DV) and (i/s/o) are always state trim.

4.9.5 *Electrical circuits*

An example of a class of continuous time systems for which these results are relevant are lumped linear electrical circuits (see Fig. 13).

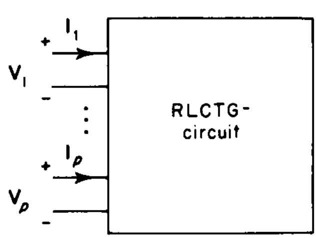


Fig. 10

Assume that the circuit contains a finite number of resistors, capacitors, inductors, transformers, gyrators, and external ports. We would like to describe the relationship imposed by the circuit on the external currents and voltages. Denote these by I_i^{ext}, V_i^{ext} with $i \in \boldsymbol{p}$ and p the number of external ports. In order to express this relationship, introduce as latent variables the current and voltages in the internal branches of the circuit. Denote these by I_i^{int}, V_i^{int} with $i \in \boldsymbol{b}$ and b the number of internal branches. Then the following equations will have to be satisfied:

(1) The *constitutive equations* in the circuit:

$$V_i = R_i I_i \qquad \text{(for a branch containing a } \textit{resistor}\text{)}$$

$$I_i = C_i \frac{dV_i}{dt} \qquad \text{(for a branch containing a } \textit{capacitor}\text{)}$$

$$V_i = L_i \frac{dI_i}{dt} \qquad \text{(for a branch containing an } \textit{inductor}\text{)}$$

$$\left.\begin{aligned} V_{i'} &= n_{i',i''} V_{i''} \\ I_{i'} &= -\frac{1}{n_{i',i''}} I_{i''} \end{aligned}\right\} \text{(for a pair of branches containing a } \textit{gyrator}\text{)}$$

$$\left.\begin{aligned} V_{i'} &= n_{i',i''} V_{i''} \\ I_{i'} &= -\frac{1}{n_{i',i''}} I_{i''} \end{aligned}\right\} \text{(for a pair of branches containing a } \textit{gyrator}\text{)}$$

(2) *Kirchhoff's current laws*: for each node (including those containing an external branch) there must hold

$$\sum_{\substack{i \\ \text{connected} \\ \text{to the node}}} I_i = 0 \quad \text{(one equation for each node)}$$

(3) *Kirchhoff's voltage laws*: For each loop (including those containing an external branch) there must hold

$$\sum_{\substack{i \\ \text{in the loop}}} V_i = 0 \quad \text{(one equation for each loop)}$$

Let $I^{ext} := \text{col}\,[I_1^{ext}, \ldots, I_p^{ext}]$, $V^{ext} := \text{col}\,[V_1^{ext}, \ldots, V_p^{ext}]$, $I^{int} := [I_1^{int}, \ldots, I_b^{int}]$, $V^{int} := \text{col}\,[V_1^{int}, \ldots, V_b^{int}]$. Organizing these equations in the obvious fashion will lead to a system of equations of the form

$$\tilde{R}_0 \begin{bmatrix} I^{ext} \\ \cdots\cdots \\ V^{ext} \end{bmatrix} + \left(\tilde{R}_1 + \tilde{R}_2 \frac{d}{dt}\right) \begin{bmatrix} I^{int} \\ \cdots\cdots \\ V^{int} \end{bmatrix} = 0$$

Note that this implies that the auxiliary variables col $[I^{int}, V^{int}]$ will define a (in general non-minimal) state variable. In fact, it is easy to see that the capacitor voltages and inductor currents form state variables. Section 4.2.3 implies that the behavioural relation between I^{ext} and V^{ext} imposed by the circuit will be described by a high order differential equation of the form

$$R'\left(\frac{d}{dt}\right) I^{ext} + R''\left(\frac{d}{dt}\right) V^{ext} = 0$$

for suitable polynomial matrices R' and R''. These equations are fully equivalent to the original system but do not involve anymore the internal currents and voltages.

4.9.6 *Recapitulation*

In this section we have studied discrete time linear invariant systems. We have assumed that the behaviour $\mathfrak{B}$ is complete. The combination of these assumptions: $T = \mathbb{Z}$, $W = \mathbb{R}^q$, and $\mathfrak{B}$ linear, shift invariant, and complete, does wonders for us: it implies, among other things, the finite dimensionality of the state space. Such systems can always be represented as the time series satisfying a finite set of autoregressive equations: linear relations among the signal variables involving linear combinations of their shifts. We have also seen that such systems always allow a (non-anticipating) input/output representation: some components of the signal variables are completely free and cause, together with the initial conditions, the remaining variables. This leads to the conclusions that these systems may be represented by means of a finite-dimensional input/state/output system having as input variables and as output variables appropriate components of the signal variables. All together this yields a clean axiomatic mathematical framework characterizing this familiar and much studied class of dynamical systems.

Important special cases of this class of systems are the controllable and the autonomous systems. These properties can be translated into properties of each of the polynomial, transfer function, or matrix representations obtained. The controllable systems are, in fact, precisely those which admit an (MA) representation: they are images of polynomial operators in the shift.

A completely analogous theory is valid for continuous time systems described by a finite set of high order linear differential equations.

4.9.7 *Sources*

The theory set forward in the section appears here for the first time in its present form. Some initial representation results have already been given in [12], [13], [14], [18], [19]. In [16] we have concentrated on continuous time system, while in [13] a number of additional results have been obtained, in particular concerning the algebraic and geometric aspects. The description of systems in terms of polynomial matrices owes much to the pioneering contributions by Rosenbrock [15] and Fuhrmann [20]. The Hautus test for controllability and observability originated in [21].

REFERENCES

[1] R. Abraham and J. E. Marsden, *Foundations of Mechanics*, Benjamin/Cummings, 1978.

[2] V. I Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, 1978.

[3] L. A. Zadeh and C. A. Desoer, *Linear System Theory—A State Space Approach*, McGraw-Hill, 1963.

[4] G. Zames, On the input-output stability of time-varying nonlinear feedback systems. Part I: Conditions derived using concepts of loop gain, conicity, and positivity, *IEEE Transactions on Automatic Control*, Vol. AC-11, pp. 228–38, 1966.

[5] J. C. Willems, *The Analysis of Feedback Systems*, MIT Press, 1971.

[6] T. G. Windeknecht, *General Dynamical Processes*, Academic Press, 1971.

[7] M. D. Mesarovic and Y. Takahara, *General Systems Theory: Mathematical Foundations*, Academic Press, 1975.

[8] S. Eilenberg, *Automata, Languages, and Machines*, Academic Press, 1974.

[9] R. E. Kalman, P. L. Falb, and M. A. Arbib, *Topics in Mathematical Systems Theory*, McGraw-Hill, 1969.

[10] R. Bellman, *Dynamic Programming*, Princeton University Press, 1957.

[11] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze and E. F. Mischenko, *The Mathematical Theory of Optimal Prcocesses*, Wiley, 1962.

[12] J. C. Willems, System theoretic models for the analysis of physical systems, *Richerche di Automatica*, Vol. 10, No. 2, pp. 71–106, 1979.

[13] A. J. Van der Schaft, *System Theoretic Description of Physical Systems*, CWI Tract No. 3, Centre for Mathematics and Computer Science, Amsterdam, 1984.

[14] J. C. Willems, From time series to linear system: [14a] Part I: Finite dimensional linear time invariant systems; [14b] Part II: Exact modelling; [14c] Part III: Approximate modelling. *Automatica*, Vol. 22, No. 5, pp. 561–80, 1986; Vol. 22, No. 6, pp. 675–94, 1986; Vol. 23, No. 1, pp. 87–115, 1987.

[15] H. H. Rosenbrock, *State Space and Multivariable Theory*, Wiley, 1970.

[16] R. E. Kalman, Irreducible realizations and the degree of a rational matrix, *SIAM Journal on Applied Mathematics*, Vol. 13, No. 2, pp. 520–44, 1965.

[17] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics, Vol. II*, Addison-Wesley, 1964.

[18] J. C. Willems, Input-output and state-space representations of finite dimensional linear time-invariant systems, *Linear Algebra and Its Applications*, Vol. 50, pp. 581–608, 1983.

[19] J. W. Nieuwenhuis and J. C. Willems, Deterministic ARMA models, *Analysis and Optimization of Systems* (Eds A. Bensoussan and J. L. Lions), Springer-Verlag Lecture Notes in Control and Information Sciences, Vol. 83, pp. 429–39, 1985.

[20] P. A. Fuhrmann, *Linear Systems and Operators in Hilbert Space*, McGraw-Hill, 1981.

[21] M. L. J. Hautus, Controllability and observability conditions of linear autonomous systems, *Proceedings Nederlandse Akademie voor Wetenschappen, Serie A*, Vol. 72, pp. 113–16, 1966.

NOTATION

Logic symbols

$\Rightarrow$	implies
$\Leftarrow$	is implied by
$\Leftrightarrow$	is equivalent to
$:\Leftrightarrow$	is by definition equivalent to
$:=$	is by definition equal to

$\exists$	there exists
$\ni$	such that
$\forall$	for all

Set theory

$\{a_1, a_2, \ldots\}$	the set consisting of the elements $a_1, a_2, \ldots$
$\{x \in X \mid x \text{ has the property } P\}$	the set of all elements of X having property P
$\varnothing$	the empty set
$\times$	Cartesian product
2^s	the set of all subsets of S

Maps

$f: M \to N; M \xrightarrow{f} N$	f maps M (domain) into N (co-domain)

A *partial map* is a map whose domain of definition may be a subset of the space on which it is defined. We will use the same notation as for maps: $f: M \to N$; $f: M \to N$ is a partial map, then $Do(f)$ denotes its domain, that is the set of points on which the action of f is defined.

$f: x \mapsto y; x \overset{f}{\mapsto} y$	f maps the element x into the element y
f^{-1}	the inverse (may be a point to set map)
$f\|_{M'}$	the restriction of f to M'
$f(M')$	the image of M' under f
$\operatorname{im} f$	the image of f
$\ker f$	the kernel of f
$f \circ g$	the composition of the maps f and g
A^B	the set of all maps from B into A
$A(\operatorname{mod} R)$	the set of all equivalence classes of A modulo R
$a(\operatorname{mod} R)$	the equivalence class associated with the element a
id_X	the identity map on X
P_a	the map (projection) from $A \times B \times \cdots$ to A defind by $P_a(a, b, \ldots) := a$
σ^t	the t-shift: $(\sigma^t f)(t') := f(t + t')$
w^-	the strict past of a time functon
w^{-0}	the past of a time function
w^+	the future of a time function
w^{+0}	the strict future of a time function
$\underset{t^-}{\Lambda}, \underset{t^+}{\Lambda}$	concatenation

Special sets

$\mathbb{N}$	the natural numbers $= \{1, 2, \ldots\}$
$\mathbb{Z}$	the integers
$\mathbb{Z}_+$	the nonnegative integers
$\mathbb{R}$	the real numbers
$\mathbb{R}_+$	the non-negative real numbers
$\mathbb{C}$	the complex numbers
$\mathbb{R}^n$	$\underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n\text{-times}}$
$\mathbb{R}^{n_1 \times n_2}$	the real matrices with n_1 rows and n_2 columns
$\mathbb{C}^n, \mathbb{C}^{n_1 \times n_2}$	analogously defined
$\mathbb{R}[s]$	the real polynomials in the indeterminate s
$\mathbb{R}[s, s^{-1}]$	the real polynomials in the indeterminates s, s^{-1}
$\mathbb{R}(s)$	the real rational functions in the indeterminate s
$\mathbb{R}_+(s)$	the proper real rational functions in the indeterminate s
$\boldsymbol{n}$	the set $\{1, 2, \ldots, n\}$

Matrices

T	transposition
$\operatorname{col}[M_1, M_2, \ldots, M_k]$	the vector or matrix $\begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_k \end{bmatrix}$
$\operatorname{diag}[a_1, a_2, \ldots, a_n]$	the (block) diagonal matrix $\begin{bmatrix} a_1 & 0 & \cdots & 0 & 0 \\ 0 & a_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & a_n \end{bmatrix}$
det	determinant
$\mathcal{G}l(n)$	$\{M \in \mathbb{R}^{n \times n} \mid \det M \neq 0\}$
rank	the rank

Function spaces

$\mathbb{L}^q$	the space of all sequences from $\mathbb{Z}$ to $\mathbb{R}^q$ equipped with the topology of pointwise convergence

$\mathscr{L}^q$	the linear shift invariant closed subspaces of $\mathbb{L}^q$
$\mathscr{L}_2(A; B)$	the square integrable maps from A to B
$\mathscr{L}^{\text{loc}}(A; B)$	the locally integrable maps from A to B
$l_1(A; B)$	the absolutely summable B-valued sequences with index in A
$l_2(A; B)$	the square summable B-valued sequences with index in A
$\mathscr{C}^k(A; B)$	the k-times continuously differentiable maps from A to B
$\mathscr{C}^0(A; B) = \mathscr{C}(A; B)$	the continuous maps from A to B
$\mathscr{C}^\infty(A; B)$	the infinitely differentiable maps from A to B