Available online at www.sciencedirect.com

ScienceDirect

Automatica 44 (2008) 1686–1696

automatica

www.elsevier.com/locate/automatica

Linear systems theory revisited[☆]

L.G. Van Willigenburg[a,*], W.L. De Koning[b,1]

[a] Systems and Control Group, Wageningen University, Technotron, P.O. Box 17, 6700 AA Wageningen, The Netherlands
[b] Delft University of Technology, The Netherlands

Received 11 October 2005; received in revised form 17 October 2007; accepted 24 October 2007
Available online 7 March 2008

Abstract

This paper investigates and clarifies how different definitions of reachability, observability, controllability, reconstructability and minimality that appear in the control literature, may be equivalent or different, depending on the type of linear system. The differences are caused by (1) whether or not the linear system has state dimensions that vary with time (2) bounds on the time axis of the linear system (3) whether or not the initial state is non-zero and (4) whether or not the system is time invariant. Also (5) time-reversibility of systems plays a role. Discrete-time linear strictly proper systems are considered. A recently published result is used to argue that all the results carry over to continuous time. Out of the investigation two types of definitions emerge. One type applies naturally to systems with constant dimensions while the other applies naturally to systems with variable dimensions. This paper reveals that time-varying (state) dimensions that are allowed to be zero are necessary to obtain equivalence between minimality and (weak) reachability together with observability at the systems level. Besides their theoretical significance the results of this paper are of practical importance for model reduction and control of time-varying discrete-time linear systems because they result in minimal realizations with smaller dimensions that are also computed more easily.

Keywords: Linear systems with time-varying dimensions; Weak reachability; Controllability; Observability; Time-reversibility; Finite horizon

1. Introduction

The last seven lines of the paper by Kalman (1962) that discovered a decomposition of continuous *time-varying* linear dynamical systems with constant dimensions, already very clearly suggest that minimal realizations of time-varying systems, in general, have *time-varying state dimensions*. Quoting Kalman, Falb, and Arbib (1969): "the only possibility of getting a reasonably well-rounded realization theory is to generalize the notion of a dynamical system in such a way that the *dimension* of the state-space is allowed to vary with time". All the more remarkable that since this discovery, systems with time-varying dimensions took a long time to be introduced. Kalman himself suggested the introduction and definition and even included a reference (Kalman et al., 1969, page 319). But we have been unable to trace the reference. In discrete time the introduction and definition seem due to Gohberg, Kaashoek, and Lerer (1992) and van der Veen and Dewilde (1992). Initially the main application area concerned computational networks such as computer algorithms where time-varying dimensions are quite natural. In continuous time the introduction of time-varying dimensions seems due to Sandberg and Rantzer (2004). This paper focused on model reduction of time-varying discrete-time linear systems. The paper also showed briefly how the discrete-time results carry over to continuous time if time-varying state dimensions are introduced. In the same manner the discrete-time results presented in this paper carry over to continuous time.

The Kalman decomposition that may be applied at arbitrary time instants immediately suggests the equivalence between minimality and reachability together with observability. An important contribution of this paper is to show that this equivalence comes out at the *systems level* only if systems with time-varying dimensions that may be *zero* are admitted.

[☆] This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Didier Henrion under the direction of Editor Roberto Tempo.

[*] Corresponding author. Tel.: +31 317482941; fax: +31 317484957.
E-mail addresses: gerard.vanwilligenburg@wur.nl (L.G. Van Willigenburg), wilros@planet.nl (W.L. De Koning).

[1] Retired from Delft University of Technology. Current address: Kroeskarper 6, 2318 NG, Leiden, The Netherlands.

0005-1098/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2007.10.021

In addition this requires considering the influence of the *free response* of systems on reachability which is also uncommon. This introduces the concept of *weak* reachability presented in this paper. Another contribution of this paper is to show how these results depend on the linear system type and also on different definitions of well-known system properties like reachability and observability.

During the development of linear systems theory, quite understandably, the definitions of the important system properties mentioned above have evolved. It is very important to be *aware* of this evolvement if one is concerned with *time-varying* linear systems. For time-invariant linear systems most of this evolvement is irrelevant because several *different* definitions lead to the same results. For example reachability and controllability are equivalent and so are observability and reconstructability for time-invariant continuous-time systems. Precisely because of this and the fact that many results (including many books) are restricted to linear time-invariant systems, there seems to be a *lack* of consistency and clarity in the definitions (of course formally, by the very nature of a definition, these do not have to be consistent at all). As this paper is very much concerned with the differences in these definitions as well as time-varying linear systems, below we describe briefly part of the history.

Kalman (1962) discovered a decomposition of linear *time-varying* continuous-time dynamical systems and showed how it is induced by the system properties called controllability and observability. These important system properties were introduced and recognised by him in the context of the regulator problem (Kalman, 1960). A little later Weiss (1964) recognised that this decomposition can actually be obtained in four different ways. His result is induced by the system properties that he called causal controllability (C-controllability) versus anti-causal controllability (A-controllability) and similarly causal observability (C-observability) versus anti-causal observability (A-observability). C-controllability coincided with controllability and C-observability coincided with observability as published in Kalman (1962). This matter was further investigated and clarified by both Weiss and Kalman (1965). It turned out to be important to distinguish between C-observability and A-observability (and dually between A-controllability and C-controllability respectively) if the system is time-varying. Given this new insight the terminology changed. A-controllability was replaced by reachability, C-controllability by controllability, A-observability by observability and C-observability by (re)constructability (Kalman et al., 1969). Reachability is *dual* to observability and controllability is *dual* to (re)constructability. This terminology seems to have settled and will be used in this paper.

Discrete-time systems, as opposed to continuous-time systems, may not be time-reversible (Weiss, 1972, with time-reversible called pointwise complete). Time-reversibility means that knowing the current state and all past inputs, all past states can be retrieved. If a system is not time-reversible then controllability to the zero state becomes a weaker property (Weiss, 1972). In this paper we will make an attempt to both simplify and unify the definitions of the system properties mentioned above such that they become *consistent* for linear time-varying systems with either constant or variable (state) dimensions. The translation of discrete-time results to continuous time in the manner described by Sandberg and Rantzer (2004) provides continuous-time systems with state dimensions that may be time variable. These continuous-time systems need no longer be time-reversible.

Our confrontation with the issues addressed in this paper occurred during our research into finite-horizon discrete-time optimal reduced-order LQG controller synthesis. In order to be minimal these controllers turned out to have time-varying dimensions (Van Willigenburg & De Koning, 1999, 2002). Moreover this type of minimality turned out not to be entirely equivalent to reachability together with observability. This triggered the investigation reported in this paper in which the equivalence is retrieved by introducing weak reachability, slightly modified definitions and zero dimensions. In this way this paper simplifies, extends and also puts into perspective the discrete-time results reported by Van Willigenburg and De Koning (1999, 2002).

The issues addressed in this paper are of practical importance for discrete time (digital) as well as continuous-time optimal reduced-order LQG controller synthesis. To maximally reduce the LQG controller dimensions time-varying dimensions are required. They are also required if actuators and sensors are not available over the full control horizon. Another important application area is model reduction of time-varying linear systems (Sandberg & Rantzer, 2004; Varga, 1999). Actually it could be argued from the results in this paper that time-varying dimensions are most natural for time-varying systems. From a theoretical point of view we became interested due to the two suggestions made by Kalman (1962) and Kalman et al. (1969). Despite these suggestions the systems and control community seems still reluctant to adopt time-varying dimensions, as we experienced from the reviews of Van Willigenburg and De Koning (1999, 2002).

The paper is organised as follows. Section 2 describes how different definitions of controllability, reconstructability, reachability, observability and minimality may be equivalent or different depending on the type of linear discrete-time system. Also it is pointed out how two different types of definitions apply naturally to systems with constant and variable dimensions respectively. Finally the equivalence between minimality and (weak) reachability together with observability is established.

Section 3 reports conclusions an important one being that time-varying (state) dimensions that are allowed to be zero should be promoted within the systems and control community.

2. Discrete-time system properties revisited

2.1. Accepting zero dimensions: advantages and consequences

Although this is considered controversial we will allow system dimensions to be zero. The reasons to allow for zero system dimensions are as follows. Firstly, zero dimensions coincide naturally with the rank of grammians being zero, which may occur in practice. Secondly, the minimal realization of systems aims at reducing the system dimensions as much as possible. This promotes to accept a zero dimension. Thirdly and most importantly, when zero dimensions are not permitted, as in Van Willigenburg and De Koning (2002), the equivalence between minimality and reachability together with observability of systems defined over a finite-time horizon is lost. This can be seen clearly from definitions 1–3 in Van Willigenburg and De Koning (2002). Another consequence of these definitions is that any time a grammian is equal to zero exceptions have to be made regarding certain results that hold at any other time. Van Willigenburg and De Koning (2002) considered explicitly the exceptions that occur at the initial and final time due to zero grammians. But the grammians may be zero at other time instants. The latter possibility was discarded. As shown in this paper, when we do accept zero dimensions, zero grammians may be identified with a zero dimension, the equivalence is retained and all the exceptions disappear. The price we pay is that we have to stretch several definitions slightly while remaining fully consistent with the original definitions when the grammians are non-zero.

Accepting zero dimensions requires accepting empty matrices and empty sequences. The following empty matrix concept is adopted where P and Q denote empty matrices,

$$P = [] \in R^{n\times 0}, \qquad Q = [] \in R^{0\times m},$$
$$PQ = 0 \in R^{n\times m}, \tag{2.1}$$
$$\text{rank}(P) = \text{rank}(Q) = \text{rank}(PQ) = 0.$$

In Eq. (2.1) R denotes the real numbers and n, m denote non-negative row and column dimensions respectively. In addition the real empty matrix,

$$QQ^T = [] \in R^{0\times 0}, \tag{2.2}$$

is considered to be *symmetric, positive definite i.e.* >0, *non-singular* and *full rank.*

Finally the following equivalence of notations applies,

$$P = [] \in R^{n\times m} \Leftrightarrow P = 0 \in R^{n\times m}, \quad n = 0 \text{ and/or } m = 0. \tag{2.3}$$

2.2. Different types of discrete-time systems

Consider the following deterministic time-varying discrete-time linear strictly proper system,

$$x_{i+1} = \Phi_i x_i + \Gamma_i u_i, \quad x_i \in R^{n_i\times 1}, u_i \in R^{m_i\times 1}, \tag{2.4}$$
$$y_i = C_i x_i, \quad y_i \in R^{l_i\times 1}. \tag{2.5}$$

In Eqs. (2.4), (2.5) i denotes the discrete time and the *non-negative* column dimensions n_i, m_i, l_i of the state x_i, the input u_i and the output y_i respectively, *may vary with time but are all assumed to be bounded*. Furthermore Φ_i, Γ_i, C_i are real matrices with compatible dimensions. In this paper we will consider and distinguish between different types of discrete-time systems. These differences relate to the time domain.

Definition 2.1. We will distinguish discrete-time systems (2.4), (2.5) the time domain of which is specified by $i \in (-\infty, \infty)$, $i \in [0, \infty)$ and $i \in [0, N]$. They are denoted as $-+$ discrete-time systems, 0+ discrete-time systems and 0N discrete-time systems respectively. 0N and 0+ discrete-time systems have a known deterministic initial state x_0 that is part of the system description.

As in Van Willigenburg and De Koning (1999, 2002) the influence of the possibly non-zero initial state x_0 of 0+ and 0N systems on several system properties will be explicitly considered in this paper.

Remark 2.1. One of the contributions of this paper is to show that the definitions selected in this paper apply to $-+$, 0+ as well as 0N discrete-time systems. *The time indices occurring in these definitions should be considered restricted to the associated time domain of the system*, as indicated in Definition 2.1. E.g. for a 0N system $\exists i$ means $\exists i \in [0, N]$ and $\forall i$ means $\forall i \in [0, N]$ except when we consider the system matrices Φ_i, Γ_i, C_i. Then, obviously, $\forall i$ means $\forall i \in [0, N-1]$.

Definition 2.2. A $-+$, 0+ or 0N discrete-time system (2.4), (2.5) is called *time invariant*, denoted by TI, if $\forall i$, $\Phi_i = \Phi, \Gamma_i = \Gamma, C_i = C$ where Φ, Γ, C are real matrices with compatible dimensions. Systems that are not time invariant are called *time-varying*, denoted by TV.

Definition 2.3. A $-+$, 0+ or 0N discrete-time system (2.4), (2.5) is said to have *constant dimensions*, denoted by CD, if $\forall i, n_i = n, m_i = m, l_i = l$ where n, m, l are bounded non-negative integers. A system that is not CD has time-varying dimensions denoted by VD.

TI systems can be of the type $-+$, 0+ and 0N. According to Definition 2.2 TI systems are CD by definition.

Associated to the empty matrix concept (2.1)–(2.3) in this paper we consider the notion of an empty system.

Definition 2.4. A $-+$, 0+ or 0N discrete-time system (2.4), (2.5) is called *empty* if $\forall i, n_i = 0$.

A $-+$, 0+ or 0N discrete-time system (2.4), (2.5) that is empty as well as CD has the properties $\forall i, n_i = 0, m_i = m, l_i = l$. This implies that $\forall i$, $\Phi_i = [] \in R^{0\times 0}$, $\Gamma_i = [] \in R^{0\times m}$, $C_i = [] \in R^{l\times 0}$. According to Definition 2.2 these empty systems are TI.

Definition 2.5. Consider $-+$, 0+ and 0N discrete-time systems (2.4), (2.5). For these systems define the following *input sequences* $U_{j,k}$ where j, k are time indices of the system,

$$U_{j,k} = \{u_j, u_{j+1}, \ldots, u_{k-1}\}, \quad k > j, U_{j,k} = \{\}, k \le j. \tag{2.6}$$

Similarly define the following *output sequences* $Y_{j,k}$,

$$Y_{j,k} = \{y_j, y_{j+1}, \ldots, y_{k-1}\}, \quad k > j, Y_{j,k} = \{\}, k \le j. \tag{2.7}$$

Finally define the *state transition matrix*,

$$\Phi_{k,j} = \Phi_{k-1}\Phi_{k-2}\cdots\Phi_j, \quad k > j, \qquad \Phi_{j,j} = I \in R^{n_j \times n_j}. \tag{2.8}$$

Definition 2.6. A −+, 0+ or 0N system (2.4), (2.5) is called *time-reversible*, denoted by RE, if $\forall i, \forall j < i$ knowing (2.4), (2.5), x_i and $U_{j,i}$ is sufficient to determine x_j. In other words at each time i knowledge of the system, the current state and all the past inputs is sufficient to determine all the past states. Systems that are not time-reversible are called *time-irreversible* denoted by IR.

Concerning time-reversibility there is the following important lemma from Weiss (1972).

Lemma 2.1. *A −+, 0+, or 0N discrete-time system (2.4), (2.5) is RE if and only if $\forall i, \forall j > i$, $\Phi_{j,i}$ is square and non-singular. In other words a discrete-time system is RE if and only if the state transition matrix is square and non-singular for any time transition.*

Lemma 2.1 reveals that any VD system that is not CD, is not RE because for some time j, $\Phi_{j+1,j} = \Phi_j$ is not square.

2.3. Controllability and reachability

Controllability is related to the ability to control the state of a system. Kalman introduced the controllability concept when he was considering the regulator problem that considers the problem of controlling any state to the zero state (Kalman, 1960). A stronger controllability concept is concerned with the possibility to control any state to any other state. It turned out later that for linear systems these two controllability concepts are equivalent only if the system is time-reversible (Weiss, 1972, with time-reversible called pointwise complete). Continuous-time systems are guaranteed to be time-reversible whereas discrete-time systems are not (Weiss, 1972). This is one reason for adopting the stronger concept. Another is that it can be linked directly to the rank of the controllability grammian of linear systems.

For the linear discrete-time system (2.4), (2.5) we have,

$$x_i = \Phi_{i,j}x_j + \sum_{k=j}^{i-1}\Phi_{i,k+1}\Gamma_k u_k, \tag{2.9}$$
$$i = j+1, j+2, \ldots, \qquad x_i = \Phi_{i,j}x_j, i = j.$$

Associated with the state transition described by Eq. (2.9) from time j to time $i \geq j$ there is the following reachability/controllability grammian,

$$W_{j,i} = \sum_{k=j}^{i-1}\Phi_{i,k+1}\Gamma_k\Gamma_k^T\Phi_{i,k+1}^T, \tag{2.10}$$
$$i = j+1, j+2, \ldots, \qquad W_{j,j} = 0.$$

This grammian is non-negative symmetric by definition. So $W_{j,i}$ non-singular is equivalent with $W_{j,i} > 0$.

Lemma 2.2. *$W_{j,k}$ non-singular ($W_{j,k} > 0$) $\Leftrightarrow \forall x_j \in R^{n_j}$, $\forall x_k \in R^{n_k}$, $\exists U_{j,k}$ that realizes the state transition $x_j \to x_k$.*

Proof. Follows immediately from Eqs. (2.9) and (2.10). □

To the next Definitions 2.7–2.11, given below, Remark 2.1 at the end of Section 2.2 applies.

Definition 2.7. A −+, 0+, or 0N systems (2.4), (2.5) is called *controllable from time* j if $\exists k(j) \geq j$ such that $\forall x_j \in R^{n_j}$, $\forall x_k \in R^{n_k}$, $\exists U_{j,k}$ that realizes the state transition $x_j \to x_k$.

Definition 2.8. A −+, 0+, or 0N systems (2.4), (2.5) is called *controllable from any time* if for $\forall j$, $\exists k(j) \geq j$ such that $\forall x_j \in R^{n_j}$, $\forall x_k \in R^{n_k}$, $\exists U_{j,k}$ that realizes the state transition $x_j \to x_k$.

Definition 2.9. A −+, 0+, or 0N systems (2.4), (2.5) is called *reachable at time* k if $\exists j(k) \leq k$ such that $\forall x_j \in R^{n_j}$, $\forall x_k \in R^{n_k}$, $\exists U_{j,k}$ that realizes the state transition $x_j \to x_k$.

Definition 2.10. A −+, 0+, or 0N system (2.4), (2.5) is called *reachable at any time* if $\forall k$, $\exists j(k) \leq k$ such that $\forall x_j \in R^{n_j}$, $\forall x_k \in R^{n_k}$, $\exists U_{j,k}$ that realizes the state transition $x_j \to x_k$.

The *only* difference between 'controllability from' and 'reachability at' is that the former is concerned exclusively with *the future* while the latter is concerned exclusively with *the past*. From Definitions 2.7–2.10 and Lemma 2.2 observe that $W_{j,k} > 0$ implies that the system is both reachable at time k and controllable from time j. Therefore the grammian (2.10) is referred to as the reachability/controllability grammian.

Definition 2.11. A −+, 0+, or 0N system (2.4), (2.5) is called *controllable from some time* or *reachable at some time* if $\exists j, k,\ j \leq k$ such that $\forall x_j \in R^{n_j}$, $\forall x_k \in R^{n_k}$, $\exists U_{j,k}$ that realizes the state transition $x_j \to x_k$.

Clearly Definitions 2.8 and 2.10 are much stronger than Definition 2.11.

Remark 2.2. A difference with the existing definitions is the appearance of the $\geq$ and $\leq$ signs instead of the usual $>$ and $<$ signs in Definitions 2.7–2.11. For 0N and 0+ systems these enable the system to be reachable at the initial time zero, namely when $n_0 = 0$. This result also requires the empty matrix concept (2.1)–(2.3) and the possibly empty sequences (2.6) and (2.7). Similarly for 0N systems they enable the system to be observable from the terminal time N, namely if $n_N = 0$. Reasons to accept these results that may be considered controversial have been mentioned at the start of Section 2.

Example 2.1. Consider a TV, 0+ system (2.4), (2.5) with $x_0 = []$, $\Phi_0 = []$, $\Gamma_0 = 1$, $\Phi_1 = \begin{bmatrix}1\\2\end{bmatrix}$, $\Gamma_1 = \begin{bmatrix}1\\0\end{bmatrix}$, $\forall i > 1 : \Phi_i = \begin{bmatrix}1 & 3\\2 & 4\end{bmatrix}$, $\Gamma_i = \begin{bmatrix}1\\0\end{bmatrix}$ This system has variable state dimensions $n_0 = 0, n_1 = 1, \forall i > 1, n_i = 2$ and a constant input dimension $m = 1$. For this system $W_{0,0} = [] \in R^{0\times 0}$. According to Eqs. (2.2), (2.3) and (2.10) this implies that $W_{0,0} > 0$. Furthermore $W_{0,1} = 1 > 0$ and finally $\forall i$, $W_{i,i+2} = \begin{bmatrix}2 & 2\\2 & 4\end{bmatrix} > 0$ holds.

Table 2.1
Types of discrete-time systems and the associated possible types of controllability and reachability

System Type	CA: Controllable from any time	RA: Reachable at any time	CR: Controllable from some time/Reachable at some time
TI, −+	+ ⇒ RA,CR	+ ⇒ CA,CR	+ ⇒ CA,RA
TI, 0+	+ ⇒ CR	−	+
TI, 0N	−	−	+
TV, CD, −+	+ ⇒ CR	+ ⇒ CR	+
TV, CD, 0+	+ ⇒ CR	−	+
TV, CD, 0N	−	−	+
TV, VD, −+	+ ⇒ CR	+ ⇒ CR	+
TV, VD, 0+	+ ⇒ CR	+ ⇒ CR	+
TV, VD, 0N	+ ⇒ CR	+ ⇒ CR	+

According to Lemma 2.2 and Definition 2.10 this implies that the system is reachable at any time.

Irrespective of the system type the conditions in Definition 2.8 as well as those in Definition 2.10 imply the conditions in Definition 2.11. Controllability from any time as well as reachability at any time clearly imply controllability from some time as well as reachability at some time. The following two examples illustrate that reachability at any time and controllability from any time are properties that do not necessarily imply one another.

Example 2.2. Consider a −+ or 0+ system (2.4), (2.5) with,

$$\begin{aligned}\Phi_i &= \begin{bmatrix}0.1 & 0.1\\ 0 & 1\end{bmatrix}, \qquad \Gamma_i = \begin{bmatrix}1\\0\end{bmatrix}, \quad i \le i_0, i_0 > 0,\\ \Phi_i &= \begin{bmatrix}0.1 & 0.1\\ 0.1 & 1\end{bmatrix}, \qquad \Gamma_i = \begin{bmatrix}1\\0\end{bmatrix}, \quad i > i_0.\end{aligned} \tag{2.11}$$

Observe that the system is not reachable at times $i \le i_0$ so it is not reachable at any time. However the system is controllable from any time. To see the latter one has to select $k > i_0$ in Definition 2.8.

Example 2.3. Consider a −+ or 0+ system (2.4), (2.5) with,

$$\begin{aligned}\Phi_i &= \begin{bmatrix}0.1 & 0.1\\ 0.1 & 0.2\end{bmatrix}, \qquad \Gamma_i = \begin{bmatrix}1\\0\end{bmatrix}, \quad i \le i_0, i_0 > 0,\\ \Phi_i &= \begin{bmatrix}0.1 & 0.1\\ 0 & 0.2\end{bmatrix}, \qquad \Gamma_i = \begin{bmatrix}1\\0\end{bmatrix}, \quad i > i_0.\end{aligned} \tag{2.12}$$

Observe that the system is not controllable from times $i > i_0$ so it is not controllable from any time while it is reachable at any time. To see the latter one has to select $k < i_0$ in Definition 2.10.

Theorem 2.1. (1) *A TI, −+ system is controllable from some time (reachable at some time) ⇔ A TI, −+ system is controllable from any time ⇔ A TI, −+ system is reachable at any time.* (2) *A TI, 0+ system is reachable at any time ⇒ it is empty.* (3) *A TI, 0N system is reachable at any time ⇒ it is empty.* (4) *A TI, 0N system is controllable from any time ⇒ it is empty.*

Proof. (1) is a direct consequence of Definitions 2.8, 2.10 and 2.11, the time-invariant nature of the system and the fact that time is not lower nor upper bounded. (2) and (3) follow from Definition 2.10 which demands that the initial state x_0 is empty due to the lower bound on time. (4) is due to the upper bound on time which according to Definitions 2.8 and 2.10 demands that x_N is empty. □

Theorem 2.2. (1) *TV, CD, 0+ systems (2.4), (2.5) cannot be reachable at any time unless they are empty.* (2) *TV, CD, 0N systems (2.4), (2.5) cannot be controllable from any time nor can they be reachable at any time unless they are empty.*

Proof. The proof follows from the same arguments that were used to proof Theorem 2.1 □

Table 2.1 summarizes Theorems 2.1 and 2.2.

2.4. Reconstructability and observability

Irrespective of the linear system type, it is well known that all system properties have a *dual property* which is the same property associated with the *dual system*. The dual system of the system (2.4), (2.5) is represented by,

$$x'_{i-1} = \Phi_i^T x'_i + C_i^T u'_i, \quad x'_i \in R^{n_i}, u'_i \in R^{l_i}, \tag{2.13}$$

$$y'_i = \Gamma_i^T x'_i, \quad y'_i \in R^{m_i}, \tag{2.14}$$

where the prime is used to indicate the state, input and output of the dual system. Note that the discrete-time of this dual system is running backwards. Therefore the dual system is called *anti-causal*. The results in this section are largely dual to those in the previous section but not entirely. The reason is that duality involves the reversal of time. Therefore for 0+ systems the results are not entirely dual. For −+ and 0N systems they are.

Associated with the dual systems (2.13), (2.14) (and therefore also with the original system (2.4), (2.5)) the following grammian is defined for $i \le j$,

$$\begin{aligned} M_{i,j} &= \sum_{k=i}^{j-1} \Phi_{k,i}^T C_k^T C_k \Phi_{k,i}, \quad i = j-1, j-2, \ldots,\\ M_{j,j} &= 0.\end{aligned} \tag{2.15}$$

System properties have been linked to this grammian by considering what the implications are when this grammian is

Table 2.2
Types of discrete-time systems and the associated possible types of observability and reconstructability

System Type	RF: Reconstructable from any time	OA: Observable at any time	RO: Reconstructable from some time/Observable at some time
TI, $-+$	$+ \Rightarrow$ OA,RO	$+ \Rightarrow$ RF,RO	$+ \Rightarrow$ RF,OA
TI, 0+	$-$	$+ \Rightarrow$ RO	$+$
TI, 0N	$-$	$-$	$+$
TV, CD, $-+$	$+ \Rightarrow$ RO	$+ \Rightarrow$ RO	$+$
TV, CD, 0+	$-$	$+ \Rightarrow$ RO	$+$
TV, CD, 0N	$-$	$-$	$+$
TV, VD, $-+$	$+ \Rightarrow$ RO	$+ \Rightarrow$ RO	$+$
TV, VD, 0+	$+ \Rightarrow$ RO	$+ \Rightarrow$ RO	$+$
TV, VD, 0N	$+ \Rightarrow$ RO	$+ \Rightarrow$ RO	$+$

non-singular. With respect to the dual systems (2.13), (2.14) the implications are the same as the ones mentioned in the previous section but the interesting implications are the ones with respect to the original systems (2.4), (2.5). Using the notation (2.7) the dual version of Lemma 2.2 reads as follows.

Lemma 2.3. *$M_{j,k}$ non-singular ($M_{j,k} > 0$) $\Leftrightarrow$ knowing $Y_{j,k}$ and the system (2.4), (2.5) is sufficient to determine the state x_j.*

Definition 2.12. A system (2.4), (2.5) is called *reconstructable from time k* if $\exists j\ (k) \leq k$ such that knowing $Y_{j,k}$ and the system (2.4), (2.5) is sufficient to determine the state x_j.

Definition 2.13. A system (2.4), (2.5) is called *reconstructable from any time* if for $\forall k$, $\exists j\ (k) \leq k$ such that knowing $Y_{j,k}$ and the system (2.4), (2.5) is sufficient to determine the state x_j.

Note that, dual to 'controllability from' the property 'reconstructability from' is stronger than the usual property which concerns the reconstruction of x_k instead of x_j in Definitions 2.12 and 2.13. To see this note that if x_j can be reconstructed then, from the system Eqs. (2.4), (2.5) and $U_{j,k}$, the state x_k follows from straightforward calculations. If the system (2.4), (2.5) is not time-reversible (pointwise complete) on the other hand, reconstruction of the state x_k is *less* demanding than reconstruction of the state x_j (Weiss, 1972 with reconstructability called determinability).

Definition 2.14. A system (2.4), (2.5) is called *observable at time j* if $\exists k\,(j) \geq j$ such that knowing $Y_{j,k}$ and the system (2.4), (2.5) is sufficient to determine the state x_j.

Definition 2.15. A system (2.4), (2.5) is called *observable at any time* if $\forall j$, $\exists k\,(j) \geq j$ such that knowing $Y_{j,k}$ and the system (2.4), (2.5) is sufficient to determine the state x_j.

Definition 2.16. A system (2.4), (2.5) is called *reconstructable from some time* or *observable at some time* if $\exists j, k,\ j \leq k$ such that knowing $Y_{j,k}$ and the system (2.4), (2.5) is sufficient to determine the state x_j.

The *only* difference between reconstructability and observability is that the former is only concerned with *past* outputs while the latter is only concerned with *future* outputs. According to the equivalence stated in Lemma 2.3, Definitions 2.12–2.15 are among the strongest possible definitions. As the results are almost dual and the non-dual results are easily derived from the results in the previous section, proofs in this section are omitted and all the important results are summarized in Table 2.2.

2.5. Non-zero initial conditions, weak reachability and bounds on state dimension changes

Within the systems and control community reachable states are generally associated with the forced response of the system associated with the system input, *disregarding* the free response associated with the initial state of the system. When designing minimal optimal LQG compensators the initial state is generally non-zero and minimal realizations are *affected* by the associated free response (Van Willigenburg & De Koning, 2002). Then to link reachability properties to the minimality of systems we must explicitly consider the influence of the free response on reachability. Van Willigenburg and De Koning (2002) only introduced a modified version of the reachability grammian to achieve this. In this section the associated system property called *weak reachability* is introduced. Consider the solution of the system equations (2.4), (2.5) given by,

$$x_i = \Phi_{i,0}x_0 + \sum_{k=0}^{i-1} \Phi_{i,k+1}\Gamma_k u_k, \quad i = 1, 2, \ldots. \tag{2.16}$$

The set of states that can be reached by a 0+ or 0N system at time i depends on the deterministic initial state x_0 of the 0+ or 0N system and therefore is denoted by $R_i\,(x_0)$. If $x_0 = 0$, $R_i\,(0)$ is completely determined by the second term on the right in Eq. (2.16) that is known as the *forced response*. Through the variation of $U_{0,i} = \{u_0, u_1, \ldots, u_{i-1}\}$, the set $R_i\,(0)$ is seen to be *a linear subspace*, i.e. *a hyperplane that contains the origin*. Suppose this linear subspace has dimension $n_i^r \leq n_i$. Then this subspace can be represented by selecting appropriately n_i^r basis vectors that span the subspace. Next consider the situation $x_0 \neq 0$. At time i two situations may be distinguished. Either the first term on the right in Eq. (2.16), i.e. the *free response* $\Phi_{i,0}x_0$, is part of the linear subspace $R_i\,(0)$ or it is not. In the former case $R_i\,(x_0)$ equals $R_i\,(0)$. In the latter case $R_i\,(x_0)$

equals $R_i(0)$ *shifted away from the origin* by $\Phi_{i,0}x_0$. In that case $R_i(x_0)$ is no longer a linear subspace but a *hyperplane* with dimension n_i^r. To represent this hyperplane instead of n_i^r now $n_i^r + 1$ basis vectors have to be selected. These can be the original n_i^r basis vectors with $\Phi_{i,0}x_0$ added to them.

Definition 2.17. For a 0+ or 0N system the linear subspace $R_i^w(x_0)$ is the linear subspace spanned by n_i^r basis vectors that span $R_i(0)$ and the vector $\Phi_{i,0}x_0$. It is called the *weakly reachable subspace* of the system (2.4), (2.5) at time i because $R_i(x_0) \subseteq R_i^w(x_0)$.

Lemma 2.4. *If at time* i, $\Phi_{i,0}x_0$ *is an element of* $R_i(0)$, *for any* $j > i$, $\Phi_{j,0}x_0$ *is an element of* $R_j(0)$.

Proof. Follows immediately from Eq. (2.16). □

Remark 2.3. Let k denote the *first* time-instant for which $\Phi_{k,0}x_0$ is an element of $R_k(0)$, if such k exists. Then the following picture emerges from Definition 2.17 and Lemma 2.4. To represent $R_i(x_0)$ either $n_i^r + 1$ or n_i^r basis vectors are required. Before time k the former is the case and $R_i(x_0) \subset R_i^w(x_0)$. At and after time k the latter is the case and $R_i(x_0) = R_i^w(x_0)$. Note that $x_0 = 0 \Rightarrow k = 0 \Rightarrow \forall i, R_i^w(0) = R_i(0)$. In other words if $x_0 = 0$, the weakly reachable and reachable subspaces are identical.

Definition 2.18. A 0+ or 0N system is called *weakly reachable at time* i if the state-space R^{n_i} equals $R_i^w(x_0)$. A 0+ or 0N system is called *weakly reachable at some time* if $\exists i$ such that the state-space R^{n_i} equals $R_i^w(x_0)$. A 0+ or 0N system is called *weakly reachable at any time* if $\forall i$ the state-space R^{n_i} equals $R_i^w(x_0)$.

Example 2.4. Consider the 0+ system (2.4), (2.5) with $\Phi_i = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $\Gamma_i = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $i = 0, 1, \ldots$ and $x_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. For this system one easily computes:

$$R_0(x_0) = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}, \qquad R_0^w(x_0) = \left\{ \begin{bmatrix} \alpha \\ 0 \end{bmatrix} \mid \alpha \in R \right\},$$
$$R_1(x_0) = \left\{ \begin{bmatrix} 1 \\ u_0 \end{bmatrix} \mid u_0 \in R \right\},$$
$$R_1^w(x_0) = \left\{ \begin{bmatrix} \alpha \\ u_0 \end{bmatrix} \mid \alpha, u_0 \in R \right\},$$
$$R_2(x_0) = \left\{ \begin{bmatrix} 1 + u_0 \\ u_1 \end{bmatrix} \mid u_0, u_1 \in R \right\},$$
$$R_2^w(x_0) = \left\{ \begin{bmatrix} \alpha \\ u_1 \end{bmatrix} \mid \alpha, u_1 \in R \right\}.$$

From this it follows that $R_i^w(x_0) = R^{n_i}$ does not hold at time $i = 0$ so the system is not weakly reachable at time $i = 0$. Therefore the system is not weakly reachable at any time. $R_i^w(x_0) = R^{n_i}$ does hold at time $i = 1$ so the system is weakly reachable at time $i = 1$. It is not difficult to see that the system is actually weakly reachable at any time $i \geq 1$. $R_i(x_0) = R^{n_i}$ does not hold at times $i = 0, 1$ whereas it does hold for $i \geq 2$. Therefore the system is not reachable at times $i = 0, 1$ so the system is not reachable at any time. Note that the system is weakly reachable at time $i = 1$ but is not reachable at time $i = 1$. This difference is crucial when determining a minimal realization of the system (Van Willigenburg & De Koning, 2002).

It is interesting to see how our definitions apply to homogeneous systems i.e. systems that have no input.

Example 2.5. Consider the 0+ system (2.4), (2.5) with $\Phi_i = 0.9$, $\Gamma_i = [] \in R^{1\times 0}$, $C_i = 1$, $i = 0, 1, \ldots$ and $x_0 = 1$. One easily computes $y_i = x_i = (0.9)^i$, $i = 0, 1, \ldots$ and also $R_i(x_0) = \left\{(0.9)^i\right\}$, $R_i^w(x_0) = \{\alpha | \alpha \in R\}$, $i = 0, 1, \ldots$. As a result this system is weakly reachable at any time and reachable at no time.

According to our definitions the reachable sets $R_i(x_0)$ of a homogeneous system at every time $i = 0, 1, \ldots$ contain the state that results from the *free response* of the homogeneous system as in Example 2.5. A homogeneous system has no forced response, only a free response. Therefore common definitions of reachable sets that *disregard* the free response for any homogeneous system produce reachable sets that are empty or just contain the zero state. This expresses the fact that if there is no input one cannot reach any states *by means* of selecting inputs. *Including* the free response implies that systems without inputs have reachable sets other than just the zero state. This *complies* with the fact that their minimal realizations in general do not have zero state dimensions.

The *reachability grammian* $W_{0,i}$ associated to a 0+ or 0N system (2.4), (2.5) is given in recursive form by,

$$W_{0,i+1} = \Phi_i W_{0,i} \Phi_i^T + \Gamma_i \Gamma_i^T \in R^{n_{i+1}\times n_{i+1}}, \quad W_{0,0} = 0 \in R^{n_0\times n_0}, \quad i > 0. \tag{2.17}$$

Similarly the *weak reachability grammian* $W'_{0,i}$ is specified in recursive form by (Van Willigenburg & De Koning, 2002),

$$W'_{0,i+1} = \Phi_i W'_{0,i} \Phi_i^T + \Gamma_i \Gamma_i^T \in R^{n_{i+1}\times n_{i+1}}, \quad W'_{0,0} = x_0 x_0^T \in R^{n_0\times n_0}, \quad i > 0. \tag{2.18}$$

This next lemma states that weak reachability may be verified by calculating the rank of the weak reachability grammian like reachability can be verified by calculating the rank of the reachability grammian.

Lemma 2.5. *For a* 0+ *or 0N system* $\dim R_i(0) = \operatorname{rank}(W_{0,i}) = n_i^r, \forall i$. *If* $k > 0$ *in Remark 2.3 exists then* $\dim R_i^w(x_0) = \operatorname{rank}\left(W'_{0,i}\right) = n_i^r + 1, i = 0, 1, \ldots, k-1, \dim R_i^w(x_0) = \operatorname{rank}\left(W'_{0,i}\right) = n_i^r, i = k, k+1, \ldots$ *else* $\operatorname{rank}\left(W'_{0,i}\right) = n_i^r + 1, \forall i$.

A 0+ or 0N system (2.4), (2.5) is reachable at time $i \Leftrightarrow W_{0,i} > 0$. A 0+ or 0N system (2.4), (2.5) is reachable at some time $\Leftrightarrow \exists i, W_{0,i} > 0$. A 0+ or 0N system (2.4), (2.5) is reachable at any time $\Leftrightarrow \forall i, W_{0,i} > 0$.

A 0+ or 0N system (2.4), (2.5) is weakly reachable at time $i \Leftrightarrow W'_{0,i} > 0$. A 0+ or 0N system (2.4), (2.5) is weakly reachable at some time $\Leftrightarrow \exists i, W'_{0,i} > 0$. A 0+ or 0N system (2.4), (2.5) is weakly reachable at any time $\Leftrightarrow \forall i, W'_{0,i} > 0$.

Proof. Follows from Van Willigenburg and De Koning (2002), Definitions 2.17, 2.18, Lemmas 2.2–2.4 and Remark 2.3. □

Lemma 2.6. *The* increase *of the state dimension of a* 0+ *or 0N system* (2.4), (2.5) *that is either reachable at any time or weakly reachable at any time is bounded from above as follows:* $\forall i, n_{i+1} - n_i \le m_i$. *The same holds for the* $-+$ *systems* (2.4) *and* (2.5) *that are reachable at any time.*

Proof. For 0+ and 0N systems (2.4) and (2.5) from Eqs. (2.17) and (2.18) observe that the increase of the rank of the reachability and weak reachability grammians with every time step is m_i at maximum. Then the result follows from the last two equivalences mentioned in Lemma 2.5. If the $-+$ system (2.4), (2.5) is reachable at any time then, according to Definition 2.10 and Lemma 2.2, $\forall i, \exists k(i) \le i$, $\text{rank}\left(W_{k,i}\right) = n_i$. Then the result follows from the following recursion that is similar to (2.17) and applies to the reachability grammian (2.10) of a $-+$ system,

$$\forall i \ge k, \quad W_{k,i+1} = \Phi_i W_{k,i} \Phi_i^T + \Gamma_i \Gamma_i^T. \quad \square \tag{2.19}$$

Without proof we state the following lemma that for $-+$ and 0N systems is dual to the previous lemma.

Lemma 2.7. The decrease *of the state dimension of a* 0+, *0N or* $-+$ *system* (2.4), (2.5) *that is observable at any time is bounded from above as follows:* $\forall i, n_{i-1} - n_i \le l_i$.

Lemma 2.8. *For CD, RE systems* (2.4), (2.5) $\text{rank}\left(W_{j,i}\right)$, $i \ge j$, *is non-decreasing with* i. *Dually for CD, RE systems* (2.4), (2.5) $\text{rank}\left(M_{i,j}\right)$ *is non-increasing with* $i \le j$. *For* 0+ *and 0N CD, RE systems* $\text{rank}\left(W'_{0,i}\right)$, $i \ge 0$ *is non-decreasing with* i.

Proof. For CD, RE systems (2.4), (2.5) the state transition matrix is always square and non-singular. Then the result follows from Eqs. (2.10), (2.15) and (2.18). □

2.6. Minimality and minimal realization

Minimal realizations of linear systems have the property that their state dimension is minimal as far as the input–output behaviour is concerned. To preserve the input–output behaviour of a linear system quite often the dimension of the state may be reduced to *different* sizes at *different* time instants. In other words quite often a minimal realization has a state-dimension that is time-variable. If on the other hand, we insist that minimal realizations are systems with constant state dimensions (are of the CD type) then, to preserve the input–output behaviour, we are forced to take the *largest* of these minimal dimensions while accepting that at other time instants several states are redundant as far as the input–output behaviour is concerned. In other words a CD system is called minimal if at *some* time the state dimension is minimal. If a minimal realization is allowed to have a state dimension that is time-varying (is allowed to be of the VD type) then a system is called minimal if its state dimension is minimal at *any* time.

In this section we will formalize these results and relate them to the system properties defined so far. From this it will become clear that for CD systems the definition of system properties containing the attribute '*at some time* ($\exists i$)' apply most naturally while for VD systems the definition of system properties containing the attribute '*at any time* ($\forall i$)' apply most naturally.

Definition 2.19. $M_{0N} : U_{0,N} \mapsto Y_{0,N}$ denotes the *input–output map* of a 0N system as determined by the system Eqs. (2.4), (2.5) and the deterministic initial state x_0. $M_{0+} : U_{0,+\infty} \mapsto Y_{0,+\infty}$ denotes the input–output map of a 0+ system as determined by the system Eqs. (2.4), (2.5) and the deterministic initial state x_0. $M_{-+} : U_{-\infty,+\infty} \mapsto Y_{-\infty,+\infty}$ is the input–output map of a $-+$ system as determined by the system Eqs. (2.4) and (2.5).

The *input–output behaviour* of a $-+$, 0+, or 0N system is fully determined by its associated input–output map. Note that if two systems with state, input and output dimensions n_i, m_i, l_i and n'_i, m'_i, l'_i respectively have identical input–output maps then $\forall i, m_i = m'_i, l_i = l'_i$.

Definition 2.20. A 0N, 0+ or $-+$ system (2.4), (2.5) with state dimensions n_i is called *minimal* if no system with state dimensions n'_i and the same input–output map exists such that $\forall i, n'_i \le n_i$ and $\exists j, n'_j < n_j$. Then the system is called a *minimal realization* of its input–output map and of any other system that has the same input–output map. A 0N, 0+ or $-+$ CD system with state dimension n is called *C-minimal* if no CD system with state dimension n' and the same input output map exists such that $n' < n$. Then the CD system is called a *C-minimal realization* of its input–output map and of any other system that has the same input–output map.

The Kalman decomposition of linear time-varying discrete-time systems at every discrete time i, when it is based on the reachable and unobservable subspace at every discrete time i, in general provides a minimal realization with variable state dimensions (Gohberg et al., 1992). Interestingly this result was already more or less obtained much earlier in Evans (1972). However, the time-varying nature of the dimension of the subspaces generated by the Kalman decomposition stopped Evans (1972) and many others following him, from further analysis. The reason to stop the analysis being that time-varying dimensions did not comply with the notion of a system at that time, since a system was presumed to have constant (state) dimensions. The Kalman decomposition can be computed entirely from the reachability and observability grammians (Boley, 1984). However, to obtain a minimal realization for 0N and 0+ systems the reachability grammian should be replaced by the *weak* reachability grammian (Van Willigenburg & De Koning, 2002). The next two theorems summarize these results.

Theorem 2.3. (1) *The* $-+$ *system* (2.4), (2.5) *is minimal* $\Leftrightarrow$ *at* any *time instant* i *the system is both reachable at time* i *as well as observable at time* i $\Leftrightarrow \forall i$,

$\exists j(i), k(i), j(i) \le i \le k(i)$ *such that* $W_{j,i} > 0, M_{i,k} > 0$.

(2) *A 0+ as well as a 0N system* (2.4), (2.5) *is minimal* $\Leftrightarrow$ *at* any *time instant i the system is both weakly reachable at time i as well as observable at time i* $\Leftrightarrow \forall i$, $\exists k(i) \ge i$ *such that* $W'_{0,i} > 0, M_{i,k} > 0$.

Theorem 2.3 states the equivalence of minimality with (weak) reachability at any time together with observability at any time. The next theorems illustrate that similar equivalences are less straightforward for C-minimality.

Theorem 2.4. (1) *A* $-+$ *system* (2.4), (2.5) *is C-minimal* $\Leftrightarrow$ *for* some *time instant k the system is both reachable at time k as well as observable at time k* $\Leftrightarrow \exists j, k, l, j \le k \le l$ *such that* $W_{j,k} > 0, M_{k,l} > 0$.

(2) *A 0+ as well as a 0N system* (2.4), (2.5) *is C-minimal* $\Leftrightarrow$ *for* some *time instant k the system is both weakly reachable at time k as well as observable at time k* $\Leftrightarrow \exists j, k, 0 \le j \le k$ *such that* $W'_{0,j} > 0, M_{j,k} > 0$.

Example 2.6. Consider the 0N, CD system with $n_i = 2$, $\Phi_i = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$, $\Gamma_i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $C_i = \begin{bmatrix} 0 & 1 \end{bmatrix}$, $i = 0, 1, 2, N = 3$, $x_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. One easily computes rank $\left(W'_{0,0}\right) = 1$, rank $\left(W'_{j,0}\right) = 2$, $j = 1, 2, 3$. Furthermore rank $\left(M_{3,3}\right) = 0$, rank $\left(M_{2,3}\right) = 1$, rank $\left(M_{k,3}\right) = 2$, $k = 0, 1$. As for $i = 2$, rank $\left(W'_{i,0}\right) =$ rank $\left(M_{i,N}\right) = n_i$ the system is C-minimal. Owing to this the system is also reachable at *some* time as well as observable from *some* time namely $i = 2$. The system is not weakly reachable at *any* time because rank $\left(W'_{0,i}\right) = n_i$ does not hold at time $i = 0$. Also the system is not observable from *any* time because rank $\left(M_{i,N}\right) = n_i$ does not hold at times $i = 2, 3$. Therefore the system is also not minimal.

Example 2.7. Using the grammians $W'_{0,i}$, $M_{i,N}$, $i = 0, 1, 2, 3 = N$ computed in Example 2.6 as an input, a Kalman decomposition of the 0N, CD system is computed at every time $i = 0, 1, 2, 3 = N$ using the algorithm presented by Boley (1984). This provides the following VD minimal realization of the 0N, CD system in Example 2.6: $n_0 = 1$, $x_0 = 1$, $\Phi_0 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $\Gamma_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $C_0 = 0$, $n_1 = 2$, $\Phi_1 = \begin{bmatrix} 1 & -1 \end{bmatrix}$, $\Gamma_1 = 0$, $C_1 = \begin{bmatrix} 0 & 1 \end{bmatrix}$, $n_2 = 1$, $\Phi_2 = [] \in R^{0\times 1}$, $\Gamma_2 = [] \in R^{0\times 1}$, $C_2 = 1$, $n_3 = 0$. Since it is a minimal realization this system is minimal which may be easily verified by calculating rank $\left(W'_{0,i}\right) =$ rank $\left(M_{i,N}\right) = n_i$, $i = 0, 1, 2, 3 = N$.

Example 2.8. Consider the 0N, CD system from Example 2.6 but with $x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Then $W'_{0,i} = W_{0,i}$, $i = 0, 1, 2, 3 = N$ and one easily computes rank $\left(W'_{0,0}\right) = 0$, rank $\left(W'_{0,1}\right) = 1$, rank $\left(W'_{j,0}\right) = 2$, $j = 2, 3$. This system is (weakly) reachable at some time namely times $i = 2, 3$ because then rank $\left(W'_{i,0}\right) = n_i$ holds. Also the system is observable from some time namely times $i = 0, 1$ because then rank $\left(M_{i,N}\right) = n_i$ holds. Still the system is not C-minimal because at no time rank $\left(W'_{0,i}\right) =$ rank $\left(M_{i,N}\right) = n_i$ holds.

Note that according to Theorem 2.4 a C-minimal system is both (weakly) reachable at some time and observable at some time. From Example 2.8 observe that the reverse does not hold in general. This is because the time at which the system is (weakly) reachable may be different from the time at which the system is observable whereas minimality demands that they are the same.

For TI systems, which are CD by definition, the next theorem establishes equivalences between C-minimality, minimality and different types of reachability, weak reachability and observability.

Theorem 2.5. (1) *A TI,* $-+$ *system is C-minimal* $\Leftrightarrow$ *A TI,* $-+$ *system is reachable at some time and is observable at some time* $\Leftrightarrow$ *A TI,* $-+$ *system is reachable at any time and is observable at any time* $\Leftrightarrow$ *A TI,* $-+$ *system is minimal.* (2) *A TI, 0+ system is C-minimal* $\Leftrightarrow \exists i_1$ *such that* $\forall i \ge i_1$ *the TI, 0+ system is weakly reachable at time i and observable at time i.* (3) *A TI, 0N system is C-minimal* $\Leftrightarrow \exists i_1, i_2$ $i_1 \le i_2$ *such that the system is weakly reachable at* any *time* $i \ge i_1$ *and observable at* any *time* $i \le i_2$.

Proof. (1) From Theorem 2.4 $\exists j, k, l, j \le k \le l$ such that $W_{j,k} > 0, M_{k,l} > 0$. But since the system is time invariant and since the time is not bounded from above nor below this implies that for *any* value of k
$\exists j, l, j \le k \le l$ such that $W_{j,k} > 0, M_{k,l} > 0$.

(2) Now, since time is bounded from below, only for a sufficiently large k, $\exists j, l, j \le k \le l$ such that $W_{j,k} > 0, M_{k,l} > 0$.

(3) If in addition time is also bounded from above time k is also upper bounded. □

TI, $-+$ systems do show the same equivalence as stated in Theorem 2.3 for TV systems. This however no longer holds for TI, 0+ and TI, 0N systems due to the lower and upper bounds on time. The next theorem applies to systems with constant dimensions (CD) that are time-varying (TV) as well as time-reversible (RE).

Theorem 2.6. (1) *TV, CD, RE,* $-+$ *systems* (2.4), (2.5) *are C-minimal* $\Leftrightarrow \exists i_1, i_2, i_1 \le i_2$ *such that the systems are reachable at* any *time* $i \ge i_1$ *and observable at* any *time* $i \le i_2$.

(2) *TV, CD, RE, 0+ and TV, CD, RE, 0N systems* (2.4), (2.5) *are C-minimal* $\Leftrightarrow \exists i_1, i_2, i_1 \le i_2$ *such that the system is weakly reachable at* any *time* $i \ge i_1$ *and observable at* any *time* $i \le i_2$.

Proof. Both (1) and (2) follow from Theorem 2.4 and Lemma 2.8. □

From Theorems 2.5 and 2.6 observe that C-minimal 0+, TV, CD, RE systems have similar properties as C-minimal 0+, TI systems. Also 0N, TV, CD, RE systems have similar properties

Table 2.3
Properties of C-minimal realizations (first 6) and minimal realizations (final 3) of systems

	CR	RO	WR	RA	CA	OA	RF	WA
TI, −+	+	+	+	+	+	+	+	+
TI, 0+	−/+	+	+	−	−	−	−	-
TI, 0N	−/+	+	+	−	−	−	−	-
TV, CD, −+	+	+	+	−	−	−	−	-
TV, CD, 0+	+	+	+	−	−	−	−	-
TV, CD, 0N	+	+	+	−	−	−	−	-
TV, VD, −+	+	+	+	+	−	+	−	+
TV, VD, 0+	−/+	+	+	−/+	−	+	−	+
TV, VD, 0N	−/+	+	+	−/+	−	+	−	+

CR: Controllable from some time (Reachable at some time), RO: Reconstructable from some time (Observable at some time), WR: Weakly reachable at some time, RA: Reachable at any time, CA: Controllable from any time, OA: Observable at any time, RF: Reconstructable from any time, WA: Weakly reachable at any time.
−/+ indicates that the property does not necessarily hold if the initial state of the associated system is non-zero and that it does hold when the initial state is zero for CD systems and empty for VD systems.

as C-minimal 0N, TI systems. For TV, CD, IR systems Lemma 2.8 does not apply and nothing significantly more can be stated about minimality than Theorem 2.4. Table 2.3 summarizes most of the results presented in this section.

The equivalences stated in Theorems 2.3–2.6 as well as the remarks made at the start of this section clearly indicate that the definitions with the attribute '*at some time* $(\exists i)$' apply most naturally to CD systems whereas definitions with the attribute '*at any time* $(\forall i)$' apply most naturally to VD systems. From the Kalman decomposition at every discrete time i the equivalence between minimality and (weak) reachability together with observability appears to be *fundamental*. By slightly modifying some definitions and by permitting systems to be of the VD type with possibly zero dimensions this equivalence now also comes out at the *systems level* which in our opinion does justice to the fundamental nature of this equivalence. The slight modifications of the definitions are entirely compatible with the former ones when the dimensions are non-zero and when time is unbounded. Moreover minimal realizations of the VD type have *smaller dimensions* than those of the CD type that are partly redundant. This redundancy is reflected by the grammians being singular at times. This in turn complicates e.g. optimal reduced-order LQG controller synthesis (Van Willigenburg & De Koning, 1999, 2002). Therefore we promote to widen the definition of systems to include VD systems, as in this paper.

3. Conclusions

An important contribution of this paper is to show that there are basically two types of definitions of controllability, reachability, reconstructability, observability and minimality of systems. One type of definition is concerned only with *some* time instant while the other type is concerned with *any* time instant. This paper demonstrated that the first type of definition applies naturally to systems with *constant* dimensions while the latter type of definition, which is much *stronger*, applies naturally to systems with *variable* state dimensions. The latter type of definition in general results in controllable, (weakly) reachable, observable, reconstructable and minimal systems with *smaller* but *variable* state dimensions, compared to their counterparts that have constant dimensions that are partly *redundant*. It has also been thoroughly investigated how several definitions of both types may be equivalent or different depending on the type of linear system they are applied to. Discrete-time linear systems of different types have been considered. To all of these our definitions consistently apply. The discrete-time results may be carried over to continuous time in the manner described briefly by Sandberg and Rantzer (2004). Currently we are investigating this translation that seems to raise some additional issues relating to instantaneous changes of the continuous-time system structure. Such changes also occur in implicit systems (Bonilla & Malabre, 1991a,b, 2000). Interestingly their description differs from the one proposed to carry over the results.

At any time instant the Kalman decomposition exposes the equivalence between minimality and (weak) reachability together with observability. Another contribution of this paper is to show that this equivalence is obtained at the *systems level* only if time-varying state dimensions that may be zero are admitted. Since we believe this equivalence to be *fundamental* through this paper we promote the acceptance of time-varying system dimensions that are allowed to be zero. In the discrete-time case variable state dimensions are well accepted since around 1992 (Gohberg et al., 1992; van der Veen & Dewilde, 1992). The paper by Sandberg and Rantzer (2004) demonstrated the usefulness of variable state dimensions in continuous time. We hope that the system definitions and results presented in this paper will further convince the systems and control community of the usefulness of variable state dimensions that are also allowed to be zero as well as the usefulness of considering the influence of the free response on reachability that leads to the concept of weak reachability. Since Gödel we are aware of the fact that even within mathematics no set of definitions is definite and to obtain further progress and understanding one *may* have to alter or stretch definitions.

Since besides the Kalman decomposition all the important system properties are directly obtained from the (weak) reachability and observability grammians, we conclude that the (weak) reachability and observability grammians as well as the Kalman decomposition of linear systems lie at the heart of linear systems theory. Having concluded this, an interesting issue that we are currently investigating concerns the differences between the *four* Kalman decompositions presented by Weiss and Kalman (1965). The sub-systems obtained from them do *not* necessarily have the same state dimension unless the continuous-time system is differentially controllable as well as differentially observable. Intuitively one expects these four apparently fundamental system decompositions to provide subsystems that do have identical dimensions. The issue turns out to be connected to the question of determining "temporal" uncontrollability/ unreachability and temporal unreconstructability/unobservability (Van Willigenburg & De Koning, 2007).

References

Boley, D. (1984). Computing the Kalman decomposition: An optimal method. *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, *AC-29*(1), 51–53.

Bonilla, M., & Malabre, M. (1991a). Variable structure systems via implicit descriptions. In *Proceedings first European control conference: Vol. 1*, (pp. 403–408).

Bonilla, M., & Malabre, M. (1991b). On the control of linear systems having internal variations. *Automatica*, *39*(11), 189–1996.

Bonilla, M., & Malabre, M. (2000). More about non-square implicit descriptions for modeling and control. In *Proceedings 39th IEEE-CDC* (pp. 3642–3647).

Evans, S. (1972). Finite-dimensional realizations of discrete-time weighting patterns. *SIAM Journal on Applied Mathematics*, *22*(1), 45–66.

Gohberg, I., Kaashoek, M. A., & Lerer, L. (1992). *Operator theory: Advances and applications*: *Vol 56*. *Minimality and realization of discrete time-varying systems* (pp. 261–296). Birkhauser.

Kalman, R. E. (1960). Contributions to the theory of optimal control. *Sociedal Matemática Mexicana*, *5*, 102–119.

Kalman, R. E. (1962). Canonical structure of linear dynamical systems. *Proceedings of the National Academy of Sciences*, *48*, 596–600.

Kalman, R. E., Falb, P. L., & Arbib, M. (1969). *Topics in mathematical systems theory*. New York: Mc Graw Hill.

Sandberg, H., & Rantzer, A. (2004). Balanced truncation of linear time-varying systems. *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, *49*, 217–229.

van der Veen, A. J., & Dewilde, P. (1992). In P. Quinton, & Y Roberts (Eds.), *Time-varying systems for computational networks, Algorithms and Parallel VLSI Architectures II*. Elsevier Science Publishers.

Van Willigenburg, L. G., & De Koning, W. L. (1999). Optimal reduced-order compensators for time-varying discrete-time systems with deterministic and white parameters. *Automatica*, *35*, 129–138.

Van Willigenburg, L. G., & De Koning, W. L. (2002). Minimal and non-minimal optimal fixed-order compensators for time-varying discrete-time systems. *Automatica*, *38*(1), 157–165.

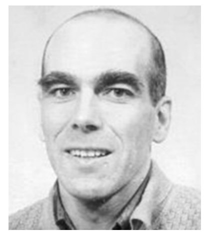
Van Willigenburg, L. G., & De Koning, W. L. (2007). A Kalman decomposition to detect temporal linear system structure. In *Proceedings European control conference* (pp. 1721–1726).

Varga, A. (1999). Balancing related methods for minimal realization of periodic systems. *Systems & Control letters*, *36*, 339–349.

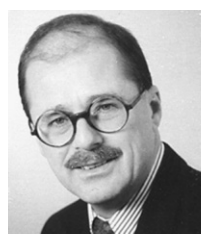
Weiss, L. (1964). Weighting patterns and the controllability and observability of linear systems. *Proceedings of National Academy of Sciences*, *51*, 1122–1127.

Weiss, L. (1972). Controllability, realization and stability of discrete-time systems. *SIAM Journal on Control*, *10*(2), 230–251.

Weiss, L., & Kalman, R. E. (1965). Contributions to linear systems theory. *International Journal of Engineering Science*, *3*, 141–171.



L.G. Van Willigenburg was born in Leiden, The Netherlands, in 1958. He received his M.Sc. in electrical engineering and his Ph.D. degree in control engineering at Delft University of Technology, The Netherlands, in 1983 and 1991 respectively. From 1986 to 1991 he was a research engineer at the process dynamics and control group in the department of applied physics. Since 1991 he is an assistant professor of the systems and control group at Wageningen University. His research interests include digital optimal control, reduced-order control, model predictive control and adaptive dual control. The application areas range from indoor climate control (storage rooms, greenhouses, stables and buildings), robot control, automatic guidance, to process control (fermentation) and the control of economic and living systems (plants).



W.L. De Koning was born in Leiden, The Netherlands, in 1944. He received his M.Sc. in electrical engineering and his Ph.D. degree in control engineering at Delft University of Technology, The Netherlands, in 1975 and 1984 respectively. From 1969 to 1975 he was a research engineer of power electronics and control at the department of electrical engineering. From 1987 to 2006 he was an associate professor of mathematical system theory at the department of mathematical system theory in the department of technical mathematics and informatics. He has also held a visiting position at the Florida Institute of Technology, Melbourne. His research interests include control of distributed parameter systems, robust control, adaptive control, reduced-order control and digital control. The main application areas are the process industry and mechatronics. Although he retired in 2006 he continues some of his research on a personal basis.