Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

A controllability test for general first-order representations

U. Helmke, J. Rosenthal and J.,M.,Schumacher

Department of Operations Reasearch, Statistics, and System Theory

BS-R9508 1995

Report BS-R9508
ISSN 0924-0659

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

A Controllability Test for General First-Order Representations

U. Helmke
Department of Mathematics, University of Regensburg
D-93040 Regensburg, Germany
HELMKE@vax1.rz.uni-regensburg.d400.de

J. Rosenthal
Department of Mathematics, University of Notre Dame,
Notre Dame, Indiana 46556, USA, and
CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
Joachim.Rosenthal@nd.edu

J. M. Schumacher
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands, and
Tilburg University, CentER and Department of Economics,
P.O. Box 90153, 5000 LE Tilburg, The Netherlands
jms@cwi.nl

Abstract

In this paper we derive a new controllability rank test for general first-order representations. The criterion generalizes the well-known controllability rank test for linear input-state systems as well as a controllability rank test by Mertzios *et al.* for descriptor systems.

AMS Subject Classification (1991): 93B05, 93C35, 93B20.
Keywords & Phrases: Controllability, multivariable linear systems, descriptor systems.
Note: The first author was partially supported by GIF grant I 184-078.6/91. The second author did this research while he was a visitor at CWI. He was supported in part by the National Science Foundation under NSF grant DMS-9400965.

1. Introduction

As is well-known (see for instance [1, 5, 14]), a general form for linear time-invariant dynamical systems is the following:

$$K\sigma x + Lx + Mw \;=\; 0 \tag{1.1}$$

where K, L, and M are matrices of sizes $(n+p)\times n$, $(n+p)\times n$, and $(n+p)\times(m+p)$ respectively. (We follow the notation of [5] and therefore denote the parameter matrices for this representation by (K, L, M) rather than (E, F, G).) Specifically, Prop. VII.3 in [14] states that a system with latent variables x and manifest variables w, over the time axis $\mathbb{Z}$, is a linear time-invariant complete state-space dynamical system if and only if it can be represented in the form (1.1) with σ denoting shift. In the continuous-time case one should interpret σ as differentiation. By various transformation algorithms it has been shown that all of the behaviors that are represented by any of the forms used in linear system theory (including matrix fraction descriptions, implicit systems, etc.) admit also a representation of the form (1.1), with appropriate identification of variables; for instance the external variable w usually denotes a vector consisting of inputs and outputs.

Properties such as observability and controllability can of course be expressed in terms of the matrices K, L, and M. In particular it is stated in [14, Prop. VII.11(v)] that (1.1) is a minimal representation of a controllable external behavior if and only if the following two conditions hold:

(i) $\lambda K + \mu L$ has full column rank for all $(\lambda, \mu) \in \mathbb{C}^2 \setminus \{(0,0)\}$;

(ii) $[\lambda K + \mu L \mid M]$ has full row rank for all $(\lambda, \mu) \in \mathbb{C}^2 \setminus \{(0,0)\}$.

Condition (i) is the observability condition whereas (ii) is the controllability condition; one readily verifies that these conditions do indeed reduce to the usual ones for the standard state space case which is obtained by taking

$$K = \begin{bmatrix} I \\ 0 \end{bmatrix}, \qquad L = \begin{bmatrix} -A \\ -C \end{bmatrix}, \qquad M = \begin{bmatrix} 0 & -B \\ I & -D \end{bmatrix}. \tag{1.2}$$

Under some circumstances one may however want to avoid the transformation to standard state space form. For instance when the entries of the matrices in the representation (1.1) are parameter dependent, different routes to arrive at the standard state space representation may have to be followed for different parameter values, so that an unattractive case by case analysis would be required. Moreover we shall below develop the controllability criterion for generalized representations that may not even be similarity equivalent to a standard state space system.

Actually there are several ways to obtain an algebraic test that is capable of deciding whether a generalized state space system of the form (1.1) is controllable. One possibility is the computation of all $(n+p) \times (n+p)$ full size minors of the pencil $[\lambda K + \mu L \mid M]$, followed by the application of a classical "multi-resultant"-test due to Macaulay [7]. A controllability rank test different from the one presented here is due to Lomadze [6] (see also [11]); this test involves a matrix of size $n(n+p) \times n(n-1+m+p)$. The distinguishing feature of the test that we shall present in this paper is that it calls for checking that a certain matrix with n rows has full row rank; moreover, the column space of this matrix can be interpreted as a reachability space (see Section 6) and for this reason we call it the *reachability matrix* of (1.1). Our test is therefore a direct generalization of the classical Kalman rank test for controllability of standard state space systems. A first step in this direction was already taken by Mertzios *et al.* [9] (see also [4]), who developed a Kalman-type test that applies to systems of the form (1.1) with $p = 0$; the present paper generalizes this work to the situation in which p is not necessarily zero.

By duality, the proposed controllability test can also be interpreted as an observability test. As such it applies to systems of the form

$$\begin{aligned} G\sigma z &= Fz \\ w &= Hz \end{aligned} \tag{1.3}$$

where F, G, and H are matrices of sizes $n \times (n+m)$, $n \times (n+m)$, and $q \times (n+m)$ respectively. We emphasize that systems of the form (1.3) (sometimes called the 'pencil form'), have the same descriptive power for smooth behaviors as the representation (1.1). The pencil form has been used recently in an investigation of 'impulsive-smooth' behaviors [2, 3], which allow solutions in a space of generalized functions. Because solutions are allowed in a larger space than usual, the resulting minimality conditions are weaker than the standard ones. In fact, the following conditions for minimality are given in [3, Thm. 4.4]:

(i) $\lambda G + \mu F$ has full row rank for some $(\lambda, \mu) \in \mathbb{C}^2 \setminus \{(0,0)\}$;

(ii) $\begin{bmatrix} \lambda G + \mu F \\ H \end{bmatrix}$ has full column rank for all $(\lambda, \mu) \in \mathbb{C}^2 \setminus \{(0,0)\}$.

Condition (ii) is the observability condition, whereas condition (i) might be called an *admissibility condition*. The set of triples (F, G, H) satisfying conditions (i) and (ii) above, considered modulo similarity equivalence, has the interesting property of being a smooth and compact projective variety [12, 6, 4, 11]. Obviously the observability condition (ii) for systems in pencil form is related by duality to the controllability condition for systems in the form (1.1), and so after simple transposition the controllability test that will be derived below can also be used to test for observability of a triple (F, G, H) in the representation (1.3).

2. Preliminaries

The purpose of this section is to review some definitions and results on adjoints of matrices. Given an $n \times n$ matrix A, the *adjoint* of A (see for instance [13, p. 7]) is the $n \times n$ matrix defined as

$$\operatorname{adj} A := \left((-1)^{i+j} \det A_{ji}\right)_{i,j=1}^{n}. \tag{2.1}$$

Here $\det A_{k\ell}$ denotes the $(n-1) \times (n-1)$ minor of A defined by omitting the k-th row and ℓ-th column from A. The adjoint of a 1×1 matrix is always 1. The following are some basic properties of the adjoint:

$$A(\operatorname{adj} A) \;=\; (\operatorname{adj} A)A \;=\; \det A \cdot I_n \tag{2.2}$$

(see [13, p. 7]);

$$\operatorname{adj}(A_1 A_2) \;=\; (\operatorname{adj} A_2)(\operatorname{adj} A_1) \tag{2.3}$$

(see [13, p. 66]). Directly from the definition we have, for scalar t,

$$\operatorname{adj}(tA) \;=\; t^{n-1} \operatorname{adj} A. \tag{2.4}$$

We now derive some lemmas that will be needed below. The first of these is actually a special case of the result of Mertzios *et al.* [9]; we include a proof that shows the relation to the standard controllability test.

Lemma 2.1 *Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n \times m}$ be given. The pair (A, B) is controllable if and only if there is no nonzero constant n-vector ξ such that $\xi^T(\operatorname{adj}(\lambda I - A))B = 0$ for all $\lambda \in \mathbb{C}$.*

Proof Write

$$(\operatorname{adj}(\lambda I - A))B \;=\; \sum_{i=1}^{n} \Gamma_i \lambda^{n-i}$$

and

$$\det(\lambda I - A) \;=\; \lambda^n + a_1 \lambda^{n-1} + \cdots + a_{n-1}.$$

Clearly, we have $\xi^T(\operatorname{adj}(\lambda I - A))B = 0$ for all λ if and only if $\xi^T\Gamma_i = 0$ for all $i = 1, \cdots, n$. Thus it remains to prove that the pair (A, B) is controllable if and only if the matrix $[\Gamma_1 \mid \cdots \mid \Gamma_n]$ has full row rank. As a consequence of (2.2), we have

$$(\lambda I - A)\sum_{i=1}^{n}\Gamma_i\lambda^{n-i} \;=\; (\det(\lambda I - A))B \;=\; \lambda^n B + a_1\lambda^{n-1}B + \cdots + a_{n-1}B.$$

By equating coefficients one obtains

$$[\Gamma_1 \mid \Gamma_2 \mid \cdots \mid \Gamma_n] \;=\; [B \mid AB \mid \cdots \mid A^{n-1}B]\begin{bmatrix} I & a_1 I & \cdots & a_{n-1}I \\ 0 & I & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_1 I \\ 0 & \cdots & 0 & I \end{bmatrix}. \tag{2.5}$$

Obviously the transformation in (2.5) is invertible, and so the matrix $[\Gamma_1 \mid \cdots \mid \Gamma_{n-1}]$ has full row rank if and only if $[B \mid \cdots \mid A^{n-1}B]$ has full row rank. But this is of course just the standard controllability test. □

The following lemma is given for matrices over a general field $\mathbb{K}$; we shall use it later in the case where $\mathbb{K} = \mathbb{R}(s)$, the field of rational functions.

LEMMA 2.2 *Let D and N be matrices over a field $\mathbb{K}$, of sizes $p \times p$ and $p \times m$ respectively. If ξ is a p-vector such that $[\xi^T \mid 0]$ belongs to the row span of $[D \mid N]$, then*

$$\xi^T(\operatorname{adj} D)N \;=\; 0. \tag{2.6}$$

If D is nonsingular then the reverse implication holds as well.

PROOF The first claim follows from the relation

$$[D \mid N]\begin{bmatrix} (\operatorname{adj} D)N \\ -(\det D)I \end{bmatrix} \;=\; 0$$

which is immediate from (2.2). If D is nonsingular, then (2.6) implies that $\xi^T D^{-1}N = 0$ and so $[\xi^T \mid 0] = \eta^T[D \mid N]$ with $\eta^T = \xi^T D^{-1}$. □

LEMMA 2.3 *Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ be given. If there exists a nonzero vector $x \in \mathbb{R}^n$ such that $x^T A = 0$ and $x^T B = 0$, then $(\operatorname{adj} A)B = 0$.*

PROOF The matrix A must be singular. If its rank is less than $n-1$, then $\operatorname{adj} A = 0$ and so certainly $(\operatorname{adj} A)B = 0$. Assume now that $\operatorname{rank} A = n-1$. Because $(\operatorname{adj} A)A = (\det A)I = 0$, all rows of $\operatorname{adj} A$ must be scalar multiples of the row vector x^T, and therefore $(\operatorname{adj} A)B = 0$. □

Let $X \in \mathbb{R}^{p\times(m+p)}$ be a matrix with more columns than rows. Let $\mathcal{I}(p, m+p)$ denote the set of all multi-indices $\alpha = (\alpha_1, \ldots, \alpha_p) \in \mathbb{N}^p$ of integers satisfying $1 \le \alpha_1 < \ldots < \alpha_p \le m+p$. For $\alpha \in \mathcal{I}(p, m+p)$ let X_α denote the $p \times p$ submatrix of X formed by the α_1-th, $\ldots, \alpha_p$-th columns of X. Let $\alpha' := \{1, \ldots, m+p\} \setminus \alpha$ denote the complementary index of α and let $X_{\alpha'}$ denote the associated $p \times m$ submatrix of X.

LEMMA 2.4 *Let* $X \in \mathbb{R}^{p\times(m+p)}$. *Then* $(\operatorname{adj} X_\alpha)X_{\alpha'} = 0$ *for all* $\alpha \in \mathcal{I}(p, m+p)$ *if and only if* $\operatorname{rank} X < p$.

PROOF To prove the necessity part let us suppose that $\operatorname{rank} X = p$. There exists $\alpha \in \mathcal{I}(p, m+p)$ such that the submatrix X_α is invertible. Without loss of generality we may assume that $\alpha = (1, \dots, p)$ and $X = [I_p \mid B]$. If $\operatorname{adj}(X_\alpha)X_{\alpha'} = 0$ for all $\alpha \in \mathcal{I}(p, m+p)$, then $B = 0$. Now consider the multi-index $\beta := (1, \dots, p-1, p+1)$. Then

$$\operatorname{adj}(X_\beta) = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

and hence $\operatorname{adj}(X_\beta)X_{\beta'} \neq 0$. Contradiction. The sufficiency part is immediate from the preceding lemma. □

3. THE CONTROLLABILITY TEST

Consider behaviors represented by

$$K\sigma x + Lx + Mw = 0 \tag{3.1}$$

where $K, L \in \mathbb{R}^{(n+p)\times n}$ and $M \in \mathbb{R}^{(n+p)\times(m+p)}$ $(n > 0, m > 0, p \geq 0)$. The system will be called 'admissible' if the rank condition

$$\operatorname{rank}(\lambda K + \mu L) = n \tag{3.2}$$

holds for some $(\lambda, \mu) \in \mathbb{C}^2$. This condition is implied by various forms of observability [14, p. 270]. Recall [14, Prop. VII.11] that a minimal (in the sense of [14]) system of the form (3.1) is *controllable* if and only if the rank condition

$$\operatorname{rank}[\lambda K + \mu L \mid M] = n + p \tag{3.3}$$

holds for all $(\lambda, \mu) \in \mathbb{C}^2 \setminus \{(0,0)\}$. Motivated by the application to develop an observability test for pencil form descriptions of impulsive-smooth behaviors, as discussed in the Introduction, we shall assume only admissibility, rather than minimality.

For $p = 0$ a Kalman-type controllability matrix for the system (3.1) was introduced by Mertzios *et al.* [9], see also Helmke [4]. It has been shown that the system is controllable if and only if the associated controllability matrix has full rank. Here we seek to extend that construction to the general case, where p is arbitrary.

For any multi-index $\alpha = (\alpha_1, \dots, \alpha_p) \in \mathcal{I}(p, m+p)$ let M_α denote the $(n+p)\times p$ submatrix of M formed by selecting the α_1-th, α_2-th, $\dots, \alpha_p$-th columns of M. Let $\alpha' := \{1, \dots, m+p\} \setminus \alpha$ denote the complementary index and let $M_{\alpha'}$ denote the associated $(n+p)\times m$ submatrix of M. Given any $(n+p)\times p$ submatrix M_α of M, write

$$\operatorname{adj}[\lambda K + \mu L \mid M_\alpha] = \begin{bmatrix} R_\alpha(\lambda,\mu) \\ S_\alpha(\lambda,\mu) \end{bmatrix} \tag{3.4}$$

where $R_\alpha(\lambda, \mu)$ and $S_\alpha(\lambda, \mu)$ are formed by the first n and last p rows of $\operatorname{adj}(\lambda K + \mu L \mid M_\alpha)$ respectively; thus $R_\alpha(\lambda, \mu)$ has size $n \times (n+p)$ and $S_\alpha(\lambda, \mu)$ has size $p \times (n+p)$. From the identity

$$\operatorname{adj}[t\lambda K + t\mu L \mid M_\alpha] = \begin{bmatrix} t^{n-1}I & 0 \\ 0 & t^n I \end{bmatrix} \operatorname{adj}[\lambda K + \mu L \mid M_\alpha]$$

we obtain that $R_\alpha(\lambda,\mu)$ and $S_\alpha(\lambda,\mu)$ are matrices of homogeneous polynomials in (λ,μ) of degrees $n-1$ and n respectively. So in particular

$$R_\alpha(\lambda,\mu)M_{\alpha'} = \sum_{i=0}^{n-1} \Gamma_{i\alpha}\lambda^i\mu^{n-1-i}. \tag{3.5}$$

for $(n \times m)$-matrices $\Gamma_{i\alpha}$, $i = 0,\ldots,n-1$. The *reachability matrix* of (K,L,M) is defined as the matrix of size $n \times nm\binom{m+p}{p}$ obtained by putting all matrices $\Gamma_{i\alpha}$ $(i = 0,\cdots,n-1;\ \alpha \in \mathcal{I}(p,m+p))$ next to each other:

$$\mathcal{R}(K,L,M) \ := \ [(\Gamma_{0\alpha} \mid \cdots \mid \Gamma_{n-1,\alpha}) \mid \alpha \in \mathcal{I}(p,m+p)]. \tag{3.6}$$

We can now state our main result.

THEOREM 3.1 *An admissible system (K,L,M) is controllable if and only if*

$$\operatorname{rank}\mathcal{R}(K,L,M) = n.$$

4. TRANSFORMATIONS

It will be convenient in the proof of the theorem to make use of various transformations on the triple (K,L,M) that do not affect the controllability properties. We begin by studying similarity transformations. Clearly, if T and S are invertible matrices, then the triple (K,L,M) is controllable if and only if (TKS^{-1},TLS^{-1},TM) is. The effect of such transformations on the matrix $\mathcal{R}(K,L,M)$ is described as follows.

LEMMA 4.1 *Let T and S be invertible $(n+p)\times(n+p)$ and $n \times n$ matrices, respectively. Then*

$$\mathcal{R}(TKS^{-1},TLS^{-1},TM) = \det T \cdot (\det S)^{-1} \cdot S\,\mathcal{R}(K,L,M). \tag{4.1}$$

PROOF For any $\alpha \in \mathcal{I}(p,m+p)$

$$\begin{aligned} \operatorname{adj}[\lambda TKS^{-1} + \mu TLS^{-1} \mid TM_\alpha]TM_{\alpha'} &= \operatorname{adj}(T[\lambda K + \mu L \mid M_\alpha]\begin{bmatrix} S^{-1} & 0 \\ 0 & I \end{bmatrix})TM_{\alpha'} \\ &= \operatorname{adj}\begin{bmatrix} S^{-1} & 0 \\ 0 & I \end{bmatrix} \operatorname{adj}[\lambda K + \mu L \mid M_\alpha](\det T)M_{\alpha'}. \end{aligned}$$

The result follows from the identity

$$\operatorname{adj}\begin{bmatrix} S^{-1} & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} (\det S)^{-1}S & 0 \\ 0 & (\det S)^{-1}I \end{bmatrix}.$$

□

REMARK 4.2 The transformations considered above are the natural transformations of system equivalence; the matrix S corresponds to a change of basis in state space, whereas the matrix T gives an invertible linear transformation of the system equations. Even in the generalized context of impulsive-smooth behaviors, the same transformation group is obtained (see [3, Thm. 4.1]). The proof above shows that the row space generated by $\mathcal{R}(K,L,M)$ is invariant under system equivalence. The same proof shows that the matrix $\mathcal{S}(K,L,M)$ that is formed from the coefficients of $S_\alpha(\lambda,\mu)M_{\alpha'}$, in analogy with (3.6), is transformed as follows:

$$\mathcal{S}(TKS^{-1}, TLS^{-1}, TM) \;=\; (\det T)(\det S)^{-1}\mathcal{S}(K,L,M).$$

It follows that the entries of $\mathcal{S}(K,L,M)$ are determined up to one multiplicative constant, or in other words that $\mathcal{S}(K,L,M)$ is a 'projective invariant'.

Apart from the similarity transformations, we shall also use the so-called 'scaling transformations' that are defined as follows. For any invertible 2×2-matrix

$$\Omega = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in GL_2(\mathbb{R})$$

write

$$(K_\Omega, L_\Omega) := (aK + bL, cK + dL).$$

Note that these transformations actually do not involve only rescaling of time, but also rotation; in particular K and L are interchanged (corresponding to time reversal in the discrete-time interpretation) by the transformation $\Omega = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. It is immediate from the characterization (3.3) that the triple (K,L,M) is controllable if and only if (K_Ω, L_Ω, M) is. To see how the controllability matrix $\mathcal{R}(K,L,M)$ changes under the scaling transformations, note that

$$\lambda(aK+bL) + \mu(cK+dL) \;=\; (a\lambda + c\mu)K + (b\lambda + d\mu)L$$

so that the effect on $R_\alpha(\lambda,\mu)$ of replacing (K,L) by $(aK+bL, cK+dL)$ is the same as replacing (λ,μ) by $(a\lambda + c\mu, b\lambda + d\mu)$. Let us first consider what the effect of such a transformation is on a scalar homogeneous polynomial

$$p(\lambda,\mu) = \sum_{i=0}^{n-1} p_i \lambda^i \mu^{n-1-i}, \quad p_j \in \mathbb{R},$$

of degree $n-1$ in the variables λ,μ. Carrying out the transformation $(\lambda,\mu) \mapsto (a\lambda + c\mu, b\lambda + d\mu)$ results in a linear transformation of the coefficients $p_0, \cdots, p_{n-1}$. For instance, for $n = 3$,

$$\sum_{i=0}^{2} p_i (a\lambda + c\mu)^i (b\lambda + d\mu)^{2-i} \;=$$

$$= \;(d^2p_0 + cdp_1 + c^2p_2)\mu^2 + (2bdp_0 + (ad+bc)p_1 + 2acp_2)\lambda\mu + (b^2p_0 + abp_1 + a^2p_2)\lambda^2.$$

Thus the new coefficients are expressed in the old ones by

$$[\tilde p_0 \mid \tilde p_1 \mid \tilde p_2] \;=\; [p_0 \mid p_1 \mid p_2] \begin{bmatrix} d^2 & 2bd & b^2 \\ cd & ad+bc & ab \\ c^2 & 2ac & a^2 \end{bmatrix}.$$

We denote the $n \times n$ transformation matrix obtained in this way by $\tau_n(\Omega)$; so for instance it follows from the above that

$$\tau_3\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) \;=\; \begin{bmatrix} d^2 & 2bd & b^2 \\ cd & ad+bc & ab \\ c^2 & 2ac & a^2 \end{bmatrix}.$$

Since $\tau_n(\Omega)\tau_n(\Omega^{-1}) = I$, the matrices $\tau_n(\Omega)$ are nonsingular. Now consider a homogeneous polynomial matrix $\Gamma(\lambda,\mu) = \sum_{i=0}^{n-1} \Gamma_i \lambda^i \mu^{n-1-i}$ whose coefficients Γ_i have size $n \times m$. The transformation $(\lambda,\mu) \mapsto (a\lambda + c\mu, b\lambda + d\mu)$ has an entrywise effect on $\Gamma(\lambda,\mu)$ and may therefore expressed in terms of the coefficients by

$$[\tilde{\Gamma}_0 \mid \cdots \mid \tilde{\Gamma}_{n-1}] \;=\; [\Gamma_0 \mid \cdots \mid \Gamma_{n-1}](I_m \otimes \tau_n(\Omega))$$

where $\otimes$ denotes the Kronecker product. Finally this transformation applies blockwise to the matrix $\mathcal{R}(K,L,M)$, where the blocks correspond to the selections α, and so we have proved the following.

LEMMA 4.3 *For each invertible* 2×2 *matrix* Ω *there exists an invertible* $n \times n$ *matrix* $\tau_n(\Omega)$ *such that*

$$\mathcal{R}(K_\Omega, L_\Omega, M) \;=\; \mathcal{R}(K,L,M)((I_m \otimes \tau_n(\Omega)) \otimes I_{\binom{m+p}{p}}). \tag{4.2}$$

REMARK 4.4 In particular, it follows that the subspace of $\mathbb{R}^n$ spanned by the columns of $\mathcal{R}(K,L,M)$ is invariant under scaling transformations.

Finally, we note that both the property of controllability and the rank of the reachability matrix are invariant under transformations of the type $(K,L,M) \mapsto (K,L,MP)$ where P is an invertible matrix. Such transformations can be interpreted as changes of basis in the space of external variables. Actually we shall only use transformations P that are permutation matrices; these correspond to just renumbering the external variables.

5. PROOF OF THE MAIN RESULT

In this section we prove Theorem 3.1. We first show the sufficiency of the stated condition for controllability. Suppose that the generalized controllability matrix $\mathcal{R}(K,L,M)$ has rank n. This immediately implies that the matrix $[\lambda K + \mu L \mid M]$ must have full row rank for some $(\lambda,\mu) \neq (0,0)$, because otherwise it follows from Lemma 2.4 that $(\mathrm{adj}\,[\lambda K + \mu L \mid M_\alpha])M_{\alpha'} = 0$ for all (λ,μ) and all α so that $\mathcal{R}(K,L,M)$ is identically zero. Consequently, $[\lambda K + \mu L \mid M]$ has full row rank for almost all (λ,μ). By assumption, we also have that $\lambda K + \mu L$ has full column rank for almost all (λ,μ), so that certainly there will be points (λ,μ) where $[\lambda K + \mu L \mid M]$ and $\lambda K + \mu L$ both have full rank. Because of the fact that both controllability and the rank of $\mathcal{R}(K,L,M)$ are invariant under scaling transformations, we may assume that this happens at $(\lambda,\mu) = (1,0)$ so that in this case K has full column rank and $[K \mid M]$ has full row rank. Permuting the columns of M if necessary, we may write $M = [M_1 \mid M_2]$ in such a way that the matrix $[K \mid M_1]$ is invertible. Now using the invariance under similarity action (from the left), we may left multiply by the inverse of $[K \mid M_1]$ and end up with K, L, and M in the following form:

$$K \;=\; \begin{bmatrix} I \\ 0 \end{bmatrix}, \qquad L \;=\; \begin{bmatrix} -A \\ -C \end{bmatrix}, \qquad M \;=\; \begin{bmatrix} 0 & -B \\ I & -D \end{bmatrix}. \tag{5.1}$$

Clearly, the matrix $[\lambda K + \mu L \mid M]$ has full row rank for all $(\lambda,\mu) \neq (0,0)$ if and only if the matrix $[\lambda I - A \mid B]$ has full row rank for all $\lambda \in \mathbb{C}$, that is to say, if and only if the pair (A,B) is controllable.

Assume now that (A,B) is *not* controllable; we want to prove that in this case the matrix $\mathcal{R}(K,L,M)$ cannot have full row rank. By Lemma 2.1 there exists a nonzero vector ξ such that $\xi^T(\mathrm{adj}\,(\lambda I - A))B = 0$. It follows from Lemma 2.2 that there exists a vector $g(\lambda)$ such that $[\xi^T \mid 0] = g^T(\lambda)[\lambda I - A \mid B]$. By the special form of the matrices K, L, and M, this implies that

$$[\xi^T \mid 0 \mid 0] \;=\; [g^T(\lambda) \mid 0][\lambda K + L \mid M_\alpha \mid M_{\alpha'}]$$

for all selections α. It follows from Lemma 2.2 that

$$[\xi^T \mid 0](\operatorname{adj}[\lambda K + L \mid M_\alpha])M_{\alpha'} \ = \ 0$$

and consequently

$$\xi^T \sum_{i=0}^{n-1} \Gamma_{i\alpha}\lambda^i \ = \ 0$$

for all λ and α. This implies that $\xi^T \Gamma_{i\alpha} = 0$ for all i and all α, so that $\xi^T \mathcal{R}(K, L, M) = 0$.

For the necessity part of the proof, we have to show that the generalized controllability matrix $\mathcal{R}(K, L, M)$ has full row rank if the matrix $[\lambda K + \mu L \mid M]$ has full rank for all $(\lambda, \mu) \neq (0, 0)$. By a suitable scaling transformation, we may assume that K, L, and M are in the form (5.1); the full rank condition then implies that the pair (A, B) is controllable. We now choose a particular selection α, namely the one for which

$$M_\alpha \ = \ \begin{bmatrix} 0 \\ I \end{bmatrix}, \qquad M_{\alpha'} \ = \ \begin{bmatrix} -B \\ -D \end{bmatrix}.$$

After some calculation, we find

$$(\operatorname{adj}[\lambda K + L \mid M_\alpha])M_{\alpha'} \ = \ -\begin{bmatrix} (\operatorname{adj}(\lambda I - A))B \\ C(\operatorname{adj}(\lambda I - A))B + (\det(\lambda I - A))D \end{bmatrix}. \tag{5.2}$$

If $\mathcal{R}(K, L, M)$ were not of full row rank, then there would exist a nonzero constant vector ξ such that $\xi^T \mathcal{R}(K, L, M) = 0$. From the above, this would imply in particular that $\xi^T(\operatorname{adj}(\lambda I - A))B = 0$. But we know from Lemma 2.1 that this contradicts the controllability of (A, B). The proof is complete.

REMARK 5.1 If $p = 0$, then the set $\mathcal{I}(p, m{+}p)$ contains just one element, and the controllability matrix $\mathcal{R}(K, L, M)$ can be written as $[\Gamma_0 \mid \cdots \mid \Gamma_{n-1}]$ where the Γ_i are the coefficients of $(\operatorname{adj}(\lambda K + \mu L))M$. This is the controllability test of Mertzios *et al.* As we have seen in the proof of Lemma 2.1, in the 'classical' case $(K, L, M) = (I, A, B)$ this is just a similarity transformation away from Kalman's controllability criterion.

REMARK 5.2 The calculation of the coefficients of $\operatorname{adj}[\lambda K + \mu L \mid M_\alpha]$ can be carried out by an adaptation of Leverrier's algorithm due to Mertzios [8].

6. THE REACHABLE SPACE

In this section we provide a dynamic interpretation of the matrix $\mathcal{R}(K, L, M)$ which also will justify the name reachability matrix. For simplicity we will restrict ourselves to discrete-time systems, i.e. we will restrict our considerations to systems of the form

$$Kx_{t+1} + Lx_t + Mw_t = 0. \tag{6.1}$$

DEFINITION 6.1 A state vector $\hat{x}$ is said to be a reachable state if there exists a sequence of states

$$\Sigma := \{x_i \in \mathbb{R}^n \mid i \in \mathcal{Z}\}$$

having the property that

1. At most finitely many vectors $x_i \in \Sigma$ are nonzero.
2. There is a set of external variables such that (6.1) is satisfied for all $t \in \mathcal{Z}$.

3. $\hat{x} \in \Sigma$.

The set of all reachable states is denoted by $\hat{\mathcal{R}}(K, L, M)$.

The set $\hat{\mathcal{R}}(K, L, M)$ can also be characterized in the following way: $\hat{x} \in \hat{\mathcal{R}}(K, L, M)$ if and only if there is a vector polynomial

$$x(\lambda) = \sum_{i=0}^{k-1} x_i \lambda^i \in \mathbb{R}^n[\lambda]$$

having $\hat{x}$ as component and a vector polynomial

$$w(\lambda) = \sum_{i=0}^{k} w_i \lambda^i \in \mathbb{R}^{m+p}[\lambda]$$

such that

$$[\lambda K + L \mid M] \begin{bmatrix} x(\lambda) \\ w(\lambda) \end{bmatrix} = 0. \tag{6.2}$$

Note that this last equation can also be written componentwise in the form

$$\left[\begin{array}{ccccc|cccccc} K & 0 & \dots & 0 & & M & 0 & \dots & \dots & 0 \\ L & K & \ddots & \vdots & & 0 & M & \ddots & & \vdots \\ 0 & L & \ddots & 0 & & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & K & & \vdots & & \ddots & M & 0 \\ 0 & \dots & 0 & L & & 0 & \dots & \dots & 0 & M \end{array}\right] \begin{bmatrix} x_0 \\ \vdots \\ x_{k-1} \\ w_0 \\ \vdots \\ w_k \end{bmatrix} \tag{6.3}$$

The following Lemma is now a simple consequence of the description (6.2).

LEMMA 6.2 *$\hat{\mathcal{R}}(K, L, M) \subset \mathbb{R}^n$ is a linear subspace.*

From the description (6.2) it is also clear that $\hat{\mathcal{R}}(K, L, M)$ is invariant under transformation of the system equation and under change of basis in the external variables, i.e. we have for $T \in Gl_{n+p}$ and $U \in Gl_{m+p}$ that

$$\hat{\mathcal{R}}(TK, TL, TMU^{-1}) \;=\; \hat{\mathcal{R}}(K, L, M). \tag{6.4}$$

The next lemma states that $\hat{\mathcal{R}}(K, L, M)$ is also invariant under scaling transformations:

LEMMA 6.3 *For each invertible 2×2 matrix Ω one has*

$$\hat{\mathcal{R}}(K_\Omega, L_\Omega, M) \;=\; \hat{\mathcal{R}}(K, L, M). \tag{6.5}$$

PROOF For every fixed positive integer k consider the set of homogeneous polynomial vectors $x(\lambda, \mu) = \sum_{i=0}^{k-1} x_i \lambda^i \mu^{k-1-i}$ whose coefficients x_i satisfy equation (6.3) for some set of external variables w_i. The transformation $(\lambda, \mu) \mapsto (a\lambda + c\mu, b\lambda + d\mu)$ has an entrywise effect on $x(\lambda, \mu)$ and may be expressed as in the proof of Lemma 4.3 through

$$[\tilde{x}_0 \mid \cdots \mid \tilde{x}_{k-1}] \;=\; [x_0 \mid \cdots \mid x_{k-1}](\tau_k(\Omega)).$$

But this establishes the invariance. □

We are now in a position to establish the connection between the subspace $\hat{\mathcal{R}}(K,L,M)$ of reachable states and the reachability matrix $\mathcal{R}(K,L,M)$ as introduced in (3.6).

THEOREM 6.4 *The vector space $\hat{\mathcal{R}}(K,L,M)$ of reachable states is equal to the column space of the reachability matrix, i.e.*

$$\hat{\mathcal{R}}(K,L,M) \;=\; \operatorname{colsp}\left(\mathcal{R}(K,L,M)\right). \tag{6.6}$$

PROOF First note that the column space of $\mathcal{R}(K,L,M)$ is certainly invariant under permutation of the external variables. After possible transformations in the internal variables and in the scaling variables and after a possible permutation of the external variables we can therefore assume that K,L,M have the special form (5.1). One readily verifies that in this situation $\hat{\mathcal{R}}(K,L,M)$ is exactly the classical reachability space

$$\operatorname{colsp}[B \mid AB \mid \cdots \mid A^{n-1}B]. \tag{6.7}$$

From the identity (5.2) it therefore follows that $\hat{\mathcal{R}}(K,L,M) \subset \operatorname{colsp}\left(\mathcal{R}(K,L,M)\right)$. On the other hand from the sufficiency part of the main proof in Section 5 it follows that any vector in the left kernel of (6.7) is also in the left kernel of $\mathcal{R}(K,L,M)$. But this establishes the proof. □

7. AN APPLICATION FROM CODING THEORY

In this section we present an application for our results which comes from coding theory. For this reason we first extend the main results obtained so far in this paper to discrete-time linear systems defined over an arbitrary base field $\mathbb{F}$ having algebraic closure $\bar{\mathbb{F}}$.

DEFINITION 7.1 Let K, L, and M be matrices of sizes $(n+p)\times n$, $(n+p)\times n$, and $(n+p)\times(m+p)$ respectively having entries from $\mathbb{F}$. Then the discrete-time linear system

$$Kx_{t+1} + Lx_t + Mw_t \;=\; 0 \tag{7.1}$$

is called observable if

(i) $\lambda K + \mu L$ has full column rank for all $(\lambda,\mu) \in \bar{\mathbb{F}}^2 \setminus \{(0,0)\}$.

It is called controllable if

(ii) $[\lambda K + \mu L \mid M]$ has full row rank for all $(\lambda,\mu) \in \bar{\mathbb{F}}^2 \setminus \{(0,0)\}$.

Conditions (i) and (ii) extend naturally the definitions given in the introduction and they were also used in [11]. By following the same reasoning as in the proof of Theorem 3.1 and Theorem 6.4 we establish the result:

THEOREM 7.2 *An admissible system of the form* (7.1) *is controllable if and only if the reachability matrix $\mathcal{R}(K,L,M)$ has full rank. Moreover the column space of $\mathcal{R}(K,L,M)$ is equal to the set $\hat{\mathcal{R}}(K,L,M)$ of reachable states.*

In the remainder of this section we will show how the results of this paper are relevant to the design of so called convolutional codes and convolutional encoders. Since convolutional codes are easily implemented and have good error correcting properties, they are among the most widely used codes in data transmission. In the sequel we explain the most basic properties of convolutional codes from a systems theory point of view. For a detailed treatment of this active research area we refer to the excellent monograph of Piret [10].

From a systems theory point of view convolutional codes are essentially the behavior of a discrete-time linear system defined over a finite field. To be precise let $\mathbb{F} = \mathbb{F}_q$ be the Galois field with q elements and consider an $(m+p) \times m$ matrix $G(D)$ defined over the polynomial ring $\mathbb{F}[D]$ in one variable D. The matrix $G(D)$ generates a convolutional code through:

$$C := \{w(D) \mid w(D) = G(D)\ell(D)\}. \tag{7.2}$$

In this so called MA-representation, the matrix $G(D)$ is called the encoder of the convolutional code C and the set of $(m+p)$-vectors $w(D)$ describe the code words, i.e. the behavior of the system. One says the convolutional encoder is catastrophic (see [10]) if the system (7.2) is not observable, i.e. if there is an element $f \in \bar{\mathbb{F}}$ having the property that $\operatorname{rank} G(f) < m$. This is also equivalent to the condition that the full size minors of $G(D)$ have a nontrivial greatest common divisor. As the name indicates it it is most important to have non-catastrophic convolutional encoders.

Instead of studying convolutional encoders through higher order representations of the form (7.2) it is equally possible to study them in terms of first order representations of the form (7.1). For this note that the realization theory as presented in [11] is valid over any commutative base field. The design of non-catastrophic encoders is therefore equivalent to the design of triples (K, L, M) that are simultaneously observable and controllable.

Although first order representations of the exact form (7.1) have not been used so far in the coding literature we would like to remark that (7.1) exactly describes the dynamics of the so called *state diagram.* (See [10] for details.) The vector $x(t)$ describes in this representation the state of the state diagram and reachability of all possible states is certainly an important design issue.

As already mentioned earlier it is typically not so difficult to design observable triples (K, L, M). Indeed it is enough to search for matrices K, L of the form

$$[\lambda K + \mu L] \;=\; \begin{bmatrix} \lambda I - \mu A \\ \mu C \end{bmatrix}$$

where A, C form an observable pair in the traditional sense. Controllability is however an issue and in a concluding example we illustrate how Theorem 3.1 and Theorem 6.4 can successfully be applied.

Example 7.3 Consider the (binary) base field $\mathbb{F} = \mathbb{F}_2 = \{0, 1\}$ and consider the encoder

$$[\lambda K + \mu L \mid M] := \left[\begin{array}{ccc|cc} \lambda & \mu & 0 & 0 & 1 \\ \lambda+\mu & \lambda & 0 & 1 & 1 \\ \mu & 0 & \lambda & 0 & 0 \\ \lambda & \mu & \mu & 1 & 1 \end{array}\right].$$

A straightforward calculation shows that the 3×3 minors of $[\lambda K + \mu L]$ are coprime, i.e. we deal with a non-catastrophic (observable) encoder of McMillan degree $n = 3$. In order to verify controllability we use Theorem 3.1 and calculate the reachability matrix:

$$\mathcal{R}(K, L, M) \;=\; \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

Obviously this matrix has not full rank and the reachable space $\hat{\mathcal{R}}(K, L, M)$ is according to Theorem 6.4 spanned by the vectors

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

8. Conclusions

We have presented a rank test for controllability of behaviors described by equations of the form $K\sigma x + Lx + Mw = 0$; similarly the dual form leads to an observability test for systems in pencil form. The test is in the spirit of Kalman's classical controllability condition; it requires checking that a certain matrix with n rows has full row rank. Moreover, the column span of this matrix has the interpretation of a reachable space. We extended the results to discrete-time systems defined over an arbitrary base field and we illustrated the results with an application from coding theory.

References

1. J. D. Aplevich (1991). *Implicit Linear Systems.* Lect. Notes Contr. Inform. Sci. 152. Springer, Berlin.
2. A. H. W. Geerts and J. M. Schumacher. Impulsive-Smooth Behavior in Multimode Systems. Part I: State-Space and Polynomial Representations. CWI Report BS-R9431, 1994.
3. A. H. W. Geerts and J. M. Schumacher. Impulsive-Smooth Behavior in Multimode Systems. Part II: Minimality and Equivalence. CWI Report BS-R9432, 1994.
4. U. Helmke. The Cohomology of Moduli Spaces of Linear Dynamical Systems. Regensburger Mathematische Schriften 24, Fak. Math. Univ. Regensburg, 1993.
5. M. Kuijper (1994). *First-Order Representations of Linear Systems.* Birkhäuser, Boston.
6. V. G. Lomadze (1990). Finite-Dimensional Time-Invariant Linear Dynamical Systems: Algebraic Theory. *Acta Appl. Math* 19, 149–201.
7. F. S. Macaulay (1903). Some Formulae in Elimination. *Proc. London Math. Soc.* XXXV, 3–27.
8. B.G. Mertzios (1984). Leverrier's algorithm for singular systems. *IEEE Trans. Automat. Contr.* AC-29, 652–653.
9. B. G. Mertzios, M. A. Christodoulou, B. L. Syrmos and F. L. Lewis (1988). Direct Controllability and Observability Time Domain Conditions of Singular Systems. *IEEE Trans. Automat. Contr.* AC-33, 788–791.
10. Ph. Piret (1988). *Convolutional Codes, an Algebraic Approach.* MIT Press, Cambridge, MA.
11. M. S. Ravi and J. Rosenthal (1995). A general Realization Theory for Higher Order Linear Differential Equations. *Systems & Control Letters*, To appear.
12. S. A. Strømme. On Parametrized Rational Curves in Grassmann Varieties. In *Space Curves*, F. Ghione, C. Peskine and E. Sernesi (eds.), Lecture Notes in Mathematics # 1266. Springer Verlag, 1987, pp. 251–272.
13. J.H.M. Wedderburn (1934). *Lectures on Matrices.* Coll. Publ. XVII. Amer. Math. Soc., Providence, R.I.
14. J. C. Willems (1991). Paradigms and Puzzles in the Theory of Dynamical Systems. *IEEE Trans. Automat. Control* AC-36, No. 3, 259–294.