Control Theory of Descriptor Systems
Lecture Notes

Lena Scholz
TU Berlin (WS 2014/15)

February 6, 2015

CONTENTS

CHAPTER

1

INTRODUCTION

A general control system can be written in the form

$$0 = F(t, x, \dot{x}, u), \qquad x(t_0) = x_0, \tag{1.1a}$$

$$y = G(t, x, u), \tag{1.1b}$$

where $F : \mathbb{I} \times \mathbb{D}_x \times \mathbb{D}_{\dot{x}} \times \mathbb{D}_u \to \mathbb{R}^l$ and $G : \mathbb{I} \times \mathbb{D}_x \times \mathbb{D}_u \to \mathbb{R}^p$ are continuous functions, $\mathbb{D}_x, \mathbb{D}_{\dot{x}} \subseteq \mathbb{R}^n$ and $\mathbb{D}_u \subseteq \mathbb{R}^m$ are open, $x_0 \in \mathbb{R}^n$ and $\mathbb{I} = [t_0, t_f] \subset \mathbb{R}$. Equation (1.1a) is called *state equation* and (1.1b) is called *ouput equation*. The continuous differentiable function $x : \mathbb{I} \to \mathbb{R}^n$ is called the *state* of the system, $u : \mathbb{I} \to \mathbb{R}^m$ the *input* or *control* and $y : \mathbb{I} \to \mathbb{R}^p$ is the *output* of the system.

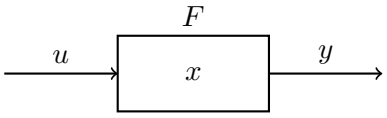


Figure 1.1: Representation of a general control system

Notation 1.1. Utilize the convenient notation

$$\dot{x}(t) = \frac{d}{dt}x(t), \qquad \ddot{x}(t) = \frac{d^2}{dt^2}x(t), \qquad \dots$$

for derivatives and

$$F_{,x} := \frac{\partial}{\partial x}F(t, x, \dot{x}, u), \qquad F_{,\dot{x}} := \frac{\partial}{\partial \dot{x}}F(t, x, \dot{x}, u)$$

for partial derivatives.

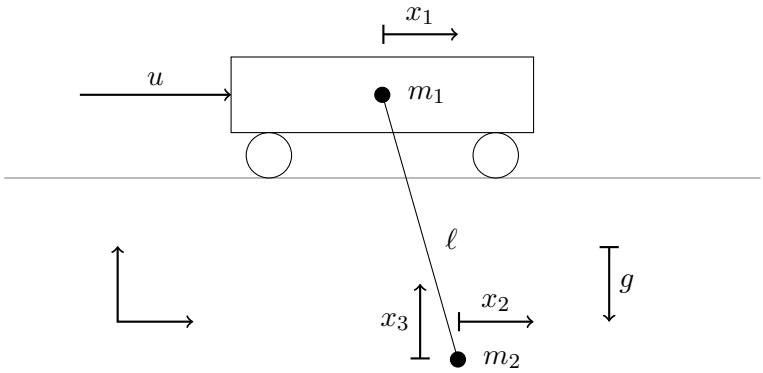


Figure 1.2: Cart pendulum

If $F_{,\dot{x}}$ is regular, the state equation (1.1a) can be reformulated as ordinary differential equation (ODE)

$$\dot{x} = \phi(t, x, u)$$

by use of the implicit function theorem. Here, we will also allow that $F_{,\dot{x}}$ is rank deficient. In this case, (1.1a) contains differential and algebraic equations, i.e., (1.1a) is a *differential-algebraic equation (DAE)*. In the control community, the system (1.1) is called a *descriptor system*. Systems of the form (1.1) arise for example in mechanical, electrical and chemical engineering.

Example 1.2 (Cart Pendulum)**.** *Consider a rigid pendulum of length ℓ with point mass m_2 attached to a cart with mass m_1 that only moves in horizontal direction. The situation is depicted in Figure 1.2.*

We have the following notation.

m_1	*mass of the cart*
m_2	*mass of the pendulum*
ℓ	*length of the pendulum*
g	*gravity*
x_1	*horizontal position of the cart*
(x_2, x_3)	*position of the mass m_2*
u	*external force acting on the car.*

The motion of the system can be described by the Euler-Lagrange equations (ELE). The Lagrange function is given by

$$L(x, \dot{x}, \lambda) = T(x, \dot{x}) - U(x) - \sum_{k=1}^{n_c} \lambda_k g_k(x),$$

where $T(x,\dot{x})$ denotes the kinetic energy, $U(x)$ denotes the potential energy and $g_1(x) = 0, \ldots, g_{n_c}(x) = 0$ denote the (ideal) constraints that restrict the motion of the system. The vector $\lambda = \begin{bmatrix}\lambda_1 & \ldots & \lambda_{n_c}\end{bmatrix}^\top$ consists of the Lagrange multipliers. *Introduce $w = \begin{bmatrix}x & \lambda\end{bmatrix}^\top$. Then the Euler-Lagrange equations are given by*

$$\frac{d}{dt}\left(\frac{\partial}{\partial \dot{w}} L(w,\dot{w})\right) - \frac{\partial}{\partial w} L(w,\dot{w}) = F_{\boldsymbol{ex}}, \tag{1.2}$$

where F_{ex} denotes some external actions (forces). In the case of the cart pendulum, we have the kinetic energy $T = \frac{1}{2}m_1\dot{x}_1^2 + \frac{1}{2}m_2\left(\dot{x}_2^2 + \dot{x}_3^2\right)$, the potential energy $U = m_2 g x_3$ and the constraint $g(x) = (x_2 - x_1)^2 + x_3^2 - \ell^2 = 0$. This yields the Lagrange function

$$L = \frac{1}{2}m_1\dot{x}_1^2 + \frac{1}{2}m_2\left(\dot{x}_2^2 + \dot{x}_3^2\right) - mgx_3 - \lambda\left((x_2 - x_1)^2 + x_3^2 - \ell^2\right).$$

Introducing $x_4 = \dot{x}_1, x_5 = \dot{x}_2$ and $x_6 = \dot{x}_3$, the ELE (1.2) is given by

$$\begin{aligned}
\dot{x}_1 &= x_4 \\
\dot{x}_2 &= x_2 \\
\dot{x}_3 &= x_3 \\
m_1\dot{x}_4 &= 2\lambda(x_2 - x_1) + u \\
m_2\dot{x}_5 &= -2\lambda(x_2 - x_1) \\
m_2\dot{x}_6 &= -2\lambda x_3 - m_2 g \\
0 &= (x_2 - x_1)^2 + x_3^2 - \ell^2.
\end{aligned} \tag{1.3}$$

Since we are only interested in the position of the pendulum, the output equation has the form

$$y = \begin{bmatrix}0 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 & 0\end{bmatrix} x = \begin{bmatrix}x_2\\ x_3\end{bmatrix},$$

where $x = \begin{bmatrix}x_1 & x_2 & \cdots & x_6 & \lambda\end{bmatrix}^\top$. Accordingly, we have $n = 7 = l$, $m = 1$ and $p = 2$.

Linearization of (1.1) along a reference trajectory leads to a *linear descriptor system with variable coefficients* of the form

$$\begin{aligned}
E(t)\dot{x}(t) &= A(t)x(t) + B(t)u(t) + f(t), \qquad x(t_0) = x_0, \\
y(t) &= C(t)x(t) + D(t)u(t) + g(t),
\end{aligned} \tag{1.4}$$

with continuous matrix functions $E, A : \mathbb{I} \to \mathbb{R}^{l\times n}, B : \mathbb{I} \to \mathbb{R}^{l\times m}, C : \mathbb{I} \to \mathbb{R}^{p\times n}$ and $D : \mathbb{I} \to \mathbb{R}^{p\times m}$ and continuous inhomogeneities $f : \mathbb{I} \to \mathbb{R}^l, g : \mathbb{I} \to \mathbb{R}^p$. Similarly, linearization of (1.1) along a constant reference trajectory leads to a *linear descriptor system with constant coefficients*

$$\begin{aligned}
E\dot{x}(t) &= Ax(t) + Bu(t) + f(t), \qquad x(t_0) = x_0, \\
y(t) &= Cx(t) + Du(t) + g(t),
\end{aligned} \tag{1.5}$$

with $E, A \in \mathbb{R}^{l\times n}, B \in \mathbb{R}^{l\times m}, C \in \mathbb{R}^{p\times n}$ and $D \in \mathbb{R}^{p\times m}$. The leading function $E(t)$ or the matrix E, respectively, are allowed to be (pointwise) singular.

Remark 1.3. In standard state-space systems (i.e. LTV or LTI) one has $E(t) = I_n = E$ and $l = n$. Hence, they are special cases of (1.4) and (1.5).

Remark 1.4 (Linearization principle)**.** Control laws based on linearization work locally for the original nonlinear system.

Topics of this course

- Solvability (consistency, regularity for descriptor systems)
- Regularization (index reduction) for example by feedback regularization
- Control theoretical aspects: C-controllability, R-controllability, I-controllability
- Stability and Stabilization
- Optimal control problems (also for nonlinear)
- Numerics: staircase form (GUPTRI), derivative array approach

Example 1.5 (Linearization of the cart pendulum)**.** *We linearize the equations of motion* (1.3) *of Example 1.2 along the equilibrium solution* $\bar{x} = \begin{bmatrix}\bar{x}_1 & \dots & \bar{x}_6, \bar{\lambda}\end{bmatrix} = \begin{bmatrix}0 & 0 & -\ell & 0 & 0 & 0, \frac{m_2 g}{2\ell}\end{bmatrix}$. *Using the decomposition* $x_i = \bar{x}_i + \hat{x}_i$ *for* $i = 1, \dots, 6$ *and* $\lambda = \bar{\lambda} + \hat{\lambda}$ *yields*

$$
\begin{aligned}
\dot{\bar{x}}_1 + \dot{\hat{x}}_1 &= \bar{x}_4 + \hat{x}_4 \\
\dot{\bar{x}}_2 + \dot{\hat{x}}_2 &= \bar{x}_5 + \hat{x}_5 \\
\dot{\bar{x}}_3 + \dot{\hat{x}}_3 &= \bar{x}_6 + \hat{x}_6 \\
m_1\left(\dot{\bar{x}}_4 + \dot{\hat{x}}_4\right) &= 2\left(\bar{x}_2 + \hat{x}_2 - \bar{x}_1 - \hat{x}_1\right)\left(\bar{\lambda} + \hat{\lambda}\right) + u \\
m_2\left(\dot{\bar{x}}_5 + \dot{\hat{x}}_5\right) &= -2\left(\bar{x}_2 + \hat{x}_2 - \bar{x}_1 - \hat{x}_1\right)\left(\bar{\lambda} + \hat{\lambda}\right) \\
m_2\left(\dot{\bar{x}}_6 + \dot{\hat{x}}_6\right) &= -2\left(\bar{x}_3 + \hat{x}_3\right)\left(\bar{\lambda} + \hat{\lambda}\right) - m_2 g \\
0 &= \left(\bar{x}_2 + \hat{x}_2 - \bar{x}_1 - \hat{x}_1\right)^2 + \left(\bar{x}_3 + \hat{x}_3\right)^2 - \ell^2.
\end{aligned}
\tag{1.6}
$$

Neglecting the quadratic terms $\hat{x}_i \hat{x}_j$ *and* $\hat{x}_i \hat{\lambda}$ *results in*

$$
\underbrace{\begin{bmatrix}
1 & & & & & & \\
& 1 & & & & & \\
& & 1 & & & & \\
& & & m_1 & & & \\
& & & & m_2 & & \\
& & & & & m_2 & \\
& & & & & & 0
\end{bmatrix}}_{E} \dot{x} = \underbrace{\begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
-\frac{m_2 g}{\ell} & \frac{m_2 g}{\ell} & 0 & 0 & 0 & 0 & 0 \\
\frac{m_2 g}{\ell} & -\frac{m_2 g}{\ell} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -\frac{m_2 g}{\ell} & 0 & 0 & 0 & 2\ell \\
0 & 0 & -2\ell & 0 & 0 & 0 & 0
\end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix}0\\0\\0\\1\\0\\0\\0\end{bmatrix}}_{B} u + \underbrace{\begin{bmatrix}0\\0\\0\\0\\0\\-m_2 g\\0\end{bmatrix}}_{f},
$$

$$
y = \underbrace{\begin{bmatrix}0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0\end{bmatrix}}_{C} x
$$

CHAPTER

2

SOLVABILITY

A first question in the analysis of descriptor systems is the existence and uniqueness of solutions of (1.1), given by

$$0 = F(t, x, \dot{x}, u), \qquad x(t_0) = x_0, \tag{2.1}$$
$$0 = y - G(t, x, u). \tag{2.2}$$

For ODEs we can employ the implicit function theorem to transform (1.1a) to

$$\dot{x} = f(t, x, u). \tag{2.3}$$

If f is a smooth function (or Lipschitz continuous with respect to the second argument) the ODE theory (for example Picard-Lindel"off) ensures a unique solution $x(t)$ for every initial condition $x_0 = x(t_0)$ and any given continuous input function u. In the general case of descriptor system this is no longer true as the following examples illustrate.

Example 2.1. *Consider the system*

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \qquad \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} x_{1,0} \\ x_{2,0} \end{bmatrix},$$

which also reads as

$$\dot{x}_2 = x_1, \qquad 0 = u.$$

We have an algebraic condition for the input u, which implies that the system is only solvable if $u \equiv 0$. On the other side, x_1 is not uniquely determined (it can be arbitrarily chosen) and can be seen as a control steering the component x_2.

Example 2.2. *The system*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{2.4}$$

consists of two differential equations for x_1 and x_2 and one algebraic relation for x_2 and u_2. The third component is not explicitly determined. Differentiating the last equation of (2.4) *and substituting into the first equation of* (2.4) *yields*

$$0 = x_2 - u_2 \quad \Longrightarrow \quad 0 = \dot{x}_2 - \dot{u}_2 = x_3 + u_1 - \dot{u}_2 \quad \Longrightarrow \quad x_3 = -u_1 + \dot{u}_2.$$

The component x_3 is implicitly defined and u_2 must be differentiable. Accordingly, system (2.4) *reads as*

$$\begin{aligned} \dot{x}_1 &= u_2 && \text{(can be stirred by } u_2\text{)}, \\ x_2 &= u_2 \\ x_3 &= -u_1 + \dot{u}_2. \end{aligned}$$

Example 2.3. *Consider the system*

$$\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

Differentiation of the first equation gives

$$\dot{x}_1 = \dot{x}_2 + \dot{u}_1.$$

Substitute this into the second equation to end up with

$$\dot{x}_2 = \frac{1}{2}\left(u_2 - \dot{u}_1\right). \tag{2.5}$$

Equation (2.5) *gives a unique solution for every initial value $x_2(t_0)$ whenever u_2 and $\dot{u}_1$ are integrable functions. The first component x_1 and its initial value are uniquely defined by the first equation.*

Definition 2.4.

1. A function $\hat{x} : \mathbb{I} \to \mathbb{R}^n$ is called a *(classical) solution* of (1.1a) if $\hat{x} \in \mathcal{C}^1(\mathbb{I}, \mathbb{R}^n)$ and $\hat{x}$ satisfies (1.1a) pointwise for some given input function u.

2. A function $\hat{x} : \mathbb{I} \to \mathbb{R}^n$ is called a *solution of the initial value problem* (IVP) consisting of (1.1a) and $x(t_0) = x_0 \in \mathbb{R}^n$, if $\hat{x}$ is a solution of (1.1a) and satisfies $\hat{x}(t_0) = x_0$.

3. An initial value $x_0 \in \mathbb{R}^n$ is called *consistent*, if the corresponding IVP has at least one solution.

Remark 2.5. There also exist weaker solvability concepts (e.g. weak solutions or impulsive smooth solutions) that can be used to handle inconsistencies ow less smoothness requirements. [later]

In some applications (e.g. robust control) it is important to know whether the system is solvable for every input function u and every initial value x_0 that is consistent with this input.

Definition 2.6. A control problem (1.1a) is called *consistent*, if there exists an input function u for which (1.1a) has a solution. It is called *regular* if it has a unique solution for every initial value that is consistent for the system with input u.

For given input u the system (1.1a) represents a differential-algebraic equation (DAE). Therefore the solvability theory for descriptor systems is strongly related to the theory for DAEs.

2.1 Linear DAEs with constant coefficients

Consider the linear DAE

$$E\dot{x} = Ax + f(t), \tag{2.6}$$

with $E, A \in \mathbb{R}^{l\times n}$ and $f : \mathbb{I} \to \mathbb{R}^l, x : \mathbb{I} \to \mathbb{R}^n$. Note that descriptor systems are a special case of (2.6) by setting $f(t) = Bu(t)$ for given input u. The solution behavior of the system depends on the properties of the matrix pair (E, A) or equivalently the matrix pencil $\lambda E - A$ for some $\lambda \in \mathbb{C}$.

Definition 2.7. A matrix pencil $\lambda E - A$ or a pair (E, A) with $E, A \in \mathbb{R}^{l\times n}$ is called *regular* if $l = n$ and $\det(\lambda E - A) \neq 0$ for some $\lambda \in \mathbb{C}$. Otherwise it is called *singular*.

Example 2.8 (Regular matrix pencil)**.** *The matrix pencil*

$$(E, A) = \left(\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

is regular, since $\det(\lambda E - A) = -1 \neq 0$ *for all* $\lambda \in \mathbb{C}$.

Example 2.9 (Singular matrix pencil)**.** *The matrix pencil*

$$(E, A) = \left(\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

is singular, since $\det(\lambda E - A) = 0$ *for all* $\lambda \in \mathbb{C}$.

Definition 2.10. Two pairs of matrices (E, A) and $(\tilde{E}, \tilde{A})$ are called *(strongly) equivalent* if there exist nonsingular matrices $W \in \mathbb{R}^{l\times l}$ and $T \in \mathbb{R}^{n\times n}$ such that

$$\tilde{E} = WET \qquad \text{and} \qquad \tilde{A} = WAT.$$

In this case, we write $(E, A) \sim (\tilde{E}, \tilde{A})$.

Lemma 2.11. *The matrix pair (E, A) is regular if and only if every strongly equivalent pair $(\tilde{E}, \tilde{A})$ is regular.*

Proof. Let $W, T \in \mathbb{R}^{n\times n}$ such that $\tilde{E} = WET$ and $\tilde{A} = WAT$. Then, we have

$$\det(\lambda\tilde{E} - \tilde{A}) = \det\left(W(\lambda E - A)T\right) = \underbrace{\det(W)\det(T)}_{\neq 0}\det(\lambda E - A),$$

which completes the proof. □

Theorem 2.12 (Weierstraß canonical form)**.** *Let $\lambda E - A$ be regular. Then there exist nonsingular matrices $W, T \in \mathbb{R}^{n\times n}$ such that*

$$\lambda WET - WAT = \lambda \begin{bmatrix} I_{n_f} & 0 \\ 0 & N \end{bmatrix} - \begin{bmatrix} J & 0 \\ 0 & I_{n_\infty} \end{bmatrix} \tag{WCF}$$

is in Weierstraß canonical form *with J, N in Jordan canonical form, N nilpotent with index of nilpotency ν, i.e. $N^\nu = 0, N^{\nu-1} \neq 0$. The number ν is called the* index *of $\lambda E - A$ or the* index *of the DAE* (2.6) *and is denoted with $\nu = ind(E, A)$.*

Proof. Since (E, A) is regular, there exists $\lambda_0 \in \mathbb{C}$ with $\det(\lambda_0 E - A) \neq 0$ and hence $\lambda_0 E - A$ is nonsingular.

$$\begin{aligned}(E, A) &= (E, A - \lambda_0 E + \lambda_0 E) \\ &\sim \left(-(\lambda_0 E - A)^{-1}E, -(\lambda_0 E - A)^{-1}(A - \lambda_0 E + \lambda_0 E)\right) \\ &= \left((A - \lambda_0 E)^{-1}E, I + \lambda_0(A - \lambda_0 E)^{-1}E\right).\end{aligned}$$

Furthermore there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that $S(A - \lambda_0 E)^{-1}S^{-1}$ is in Jordan canonical form, i.e.

$$S(A - \lambda_0 E)^{-1}S^{-1} = \begin{bmatrix} \tilde{J} & 0 \\ 0 & \tilde{N} \end{bmatrix},$$

where $\tilde{J}$ is nonsingular (part belonging to the nonzero eigenvalues) and $\tilde{N}$ is nilpotent strictly upper triangular. Then the matrix $I + \lambda_0\tilde{N}$ is nonsingular upper triangular and we have

$$\begin{aligned}(E, A) &\sim \left(\begin{bmatrix} \tilde{J} & 0 \\ 0 & \tilde{N} \end{bmatrix}, \begin{bmatrix} I + \lambda_0\tilde{J} & 0 \\ 0 & I + \lambda_0\tilde{N} \end{bmatrix}\right) \\ &\sim \left(\begin{bmatrix} I & 0 \\ 0 & (I - \lambda_0\tilde{N})^{-1}\tilde{N} \end{bmatrix}, \begin{bmatrix} \tilde{J}^{-1} + \lambda_0 I & 0 \\ 0 & I \end{bmatrix}\right),\end{aligned}$$

with $(I + \lambda_0 \tilde{N})^{-1}\tilde{N}$ is strictly upper triangular and nilpotent. Transferring $\tilde{J}^{-1} + \lambda_0 I$ and $(I + \lambda_0 \tilde{N})^{-1}\tilde{N}$ to Jordan canonical form yields the desired form (WCF). □

For regular matrix pairs (E, A) the linear DAE (2.6) can be transformed into Weierstraß canonical form by

$$\begin{aligned} & WETT^{-1}\dot{x} = WATT^{-1}x + Wf(t) \\ \iff & \begin{bmatrix} I_{n_f} & 0 \\ 0 & N \end{bmatrix} \begin{bmatrix} \dot{\tilde{x}}_1 \\ \dot{\tilde{x}}_2 \end{bmatrix} = \begin{bmatrix} J & 0 \\ 0 & I_{n_\infty} \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} + \begin{bmatrix} \tilde{f}_1 \\ \tilde{f}_2 \end{bmatrix} \end{aligned}$$

by using the variable transformation $\tilde{x} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = T^{-1}x$ and $\tilde{f} = \begin{bmatrix} \tilde{f}_1 \\ \tilde{f}_2 \end{bmatrix} = Wf$. The resulting system is decoupled into the ordinary differential equation

$$\dot{\tilde{x}}_1 = J\tilde{x}_1 + \tilde{f}_1, \tag{2.7}$$

which is called the *differential part* (also known as *dynamic part* or *slow part*) of the system (2.6) and the algebraic equation

$$N\dot{\tilde{x}}_2 = \tilde{x}_2 + \tilde{f}_2, \tag{2.8}$$

also known as the *algebraic part* or *slow subsystem* of (2.6). To see, that (2.8) is indeed an algebraic equation, we consider (2.6) in its matrix form

$$\begin{bmatrix} 0 & * & & \\ & \ddots & \ddots & \\ & & \ddots & * \\ & & & 0 \end{bmatrix} \begin{bmatrix} \dot{\tilde{x}}_{2,1} \\ \vdots \\ \vdots \\ \dot{\tilde{x}}_{2,n_\infty} \end{bmatrix} = \begin{bmatrix} \tilde{x}_{2,1} \\ \vdots \\ \vdots \\ \tilde{x}_{2,n_\infty} \end{bmatrix} + \begin{bmatrix} \tilde{f}_{2,1} \\ \vdots \\ \vdots \\ \tilde{f}_{2,n_\infty} \end{bmatrix},$$

where $* \in \{0, 1\}$. The last equation uniquely determines $\tilde{x}_{2,n_\infty} = -\tilde{f}_{2,n_\infty}$. Using this relation, one can substitute in the second last equation and solve the algebraic equation. Continuing with this procedure uniquely determines all components of $\tilde{x}_2$. In particular, the following result holds.

Lemma 2.13. *The solution $\tilde{x}_2$ of the algebraic equation* (2.8) *is given by*

$$\tilde{x}_2 = -\sum_{i=0}^{\nu-1} N^i \tilde{f}_2^{(i)}(t), \tag{2.9}$$

where ν denotes the index of nilpotency of N.

Proof. Let $D = \frac{d}{dt}$ denote the differentiation operator. Since N is constant matrix, D and N commute and also $(ND)^\nu = N^\nu D^\nu = 0$. Moreover,

$$(I - ND)\sum_{i=0}^{\nu-1} (ND)^i = I - N^\nu D^\nu = I.$$

Finally, (2.8) can be rewritten as $(I - ND)\tilde{x}_2 = -\tilde{f}_2$, which yields

$$\tilde{x}_2 = -(I - ND)^{-1}\tilde{f}_2 = -\sum_{i=0}^{\nu-1} N^i D^i \tilde{f}_2 = -\sum_{i=0}^{\nu-1} N^i \tilde{f}_2^{(i)}.$$

□

Thus, $\tilde{x}_2$ is uniquely determined by the algebraic equation (2.9). The transformed initial value $\tilde{x}_{2,0}$ has to satisfy the algebraic equation (2.9), i.e. it has to be consistent since otherwise the system is not solvable. Moreover, the inhomogeneity $\tilde{f}_2$ (and hence also f) has to be $(\nu - 1)$ times continuously differentiable. On the other side, system (2.7) has a unique solution $\tilde{x}_1$ for any initial value $\tilde{x}_{1,0}$ and any inhomogeneity $\tilde{f}_1$ given by

$$\tilde{x}_1(t) = e^{Jt}\tilde{x}_{1,0} + \int_0^t e^{J(t-s)}\tilde{f}_1(s)ds.$$

The backtransformation finally gives the solution $x = T\tilde{x}$.

Remark 2.14. In the descriptor setting the inhomogeneity $f(t)$ is given by $f(t) = Bu(t)$. Thus $\begin{bmatrix}\tilde{f}_1\\ \tilde{f}_2\end{bmatrix} = WBu(t) = \begin{bmatrix}\tilde{B}_1\\ \tilde{B}_2\end{bmatrix} u(t)$ and $\tilde{B}_2 u(t)$, i.e. $u(t)$ has to be $(\nu-1)$ times continuously differentiable. For the existence of a classical solution x we need $u \in \mathcal{C}^\nu(\mathbb{I}, \mathbb{R})$. Thus piecewise continuous control functions (or bang-bang control) might not work. Consistency of initial conditions may depend on the derivatives of the input function $u(t)$.

We summarize the previous results in the following theorem.

Theorem 2.15 (Existence and Uniqueness of solutions)**.** *Consider a linear DAE with constant coefficients* (2.6) *with regular matrix pair* (E, A) *and inhomogeneity* $f \in \mathcal{C}^\nu(\mathbb{I}, \mathbb{R}^n)$ *where* $\nu = ind(E, A)$*. Then it holds that:*

1. *The DAE* (2.6) *is solvable.*

2. *An initial value* $x_0 \in \mathbb{R}^n$ *is consistent if and only if*

$$\tilde{x}_{2,0} = -\sum_{i=1}^{\nu-1} N^i \tilde{f}_2^{(i)}(0),$$

   where $T^{-1}x_0 = \begin{bmatrix}\tilde{x}_{1,0}\\ \tilde{x}_{2,0}\end{bmatrix}$ *and* $Wf = \begin{bmatrix}\tilde{f}_1\\ \tilde{f}_2\end{bmatrix}$ *with* $T, W \in \mathbb{R}^{n\times n}$ *that transform* (E, A) *to Weierstraß canonical form* (WCF)*.*

3. *Every initial problem* (2.6) *with consistent initial value* x_0 *is uniquely solvable.*

Definition 2.16. The set of consistent initial values is defines as

$$\mathcal{X}_c^0 := \left\{ x_0 = T \begin{bmatrix} \tilde{x}_{1,0} \\ \tilde{x}_{2,0} \end{bmatrix} \mid \tilde{x}_{1,0} \in \mathbb{R}^{n_f}, \tilde{x}_{2,0} = -\sum_{i=0}^{\nu-1} N^i f^{(i)}(0) \right\}.$$

We conclude that if (E, A) is regular, $x_0 \in \mathcal{X}_c^0$ and $u(t)$ is ν times continuously differentiable, then

$$E\dot{x} = Ax + Bu, \qquad x(0) = x_0$$

has a unique (classical) solution. To clarify the dependency of the solution x on the initial value and the input, we write $x(t; x_0, u)$.

Theorem 2.17. *If the pair of matrices (E, A) is regular, then the control problem* (1.5) *is consistent and regular.*

Proof. Taking $u(t) \equiv 0$, we have $E\dot{x} = Ax + f$ with regular matrix pair (E, A). Thus, for consistent initial value $x(0) = x_0$ there exists a (unique) solution. Hence, the control problem is consistent. Regularity follows from Theorem 2.15. □

Theorem 2.18. *If (E, A) with $E, A \in \mathbb{R}^{l\times n}$ is a singular matrix pair, then the control problem* (1.5) *is not regular.*

Proof. **Case 1** $\operatorname{rank}(\lambda E - A) < n$ for all $\lambda \in \mathbb{C}$.

We choose $u \equiv 0$ and $f(t) \equiv 0$ and consider the homogeneous DAE $E\dot{x} = Ax$ together with $x(0) = x_0$. Let $\lambda_1, \ldots, \lambda_{n+1} \in \mathbb{C}$ be pairwise different. Then for every λ_i there exists $v_i \in \mathbb{C}^n \backslash \{0\}$ with

$$(\lambda_i E - A)v_i = 0$$

and the v_i are linearly dependent. Hence, there exists $\alpha_i \in \mathbb{C}$ $(i = 1, \ldots, n+1)$ not all of them being zero such that $\sum_{i=1}^{n+1} \alpha_i v_i = 0$. Define $x(t) = \sum_{i=1}^{n+1} \alpha_i v_i e^{\lambda_i t}$. Then $x(0) = 0$ and

$$E\dot{x}(t) = E\sum_{i=1}^{n+1} \alpha_i \lambda_i e^{\lambda_i t} = A\sum_{i=1}^{n+1} \alpha_i v_i e^{\lambda_i t} = Ax(t).$$

Thus, $x(t)$ is a solution of the homogeneous system with $x(0) = 0$. Since $\tilde{x}(t) \equiv 0$ is also a solution, the solution is not unique, i.e. the control problem is not regular.

Case 2 $\operatorname{rank}(\bar{\lambda}E - A) = n$ for some $\bar{\lambda} \in \mathbb{C}$.

Since (E, A) is singular, this implies $l > n$. With the variable transformation $x(t) = e^{\bar{\lambda}t}\tilde{x}(t)$ we get

$$E\left(e^{\bar{\lambda}t}\dot{\tilde{x}}(t) + \bar{\lambda}e^{\bar{\lambda}t}\tilde{x}(t)\right) = Ae^{\bar{\lambda}t}\tilde{x}(t) + Bu(t) + f(t).$$

In particular, $E\dot{\tilde{x}}(t) = (A - \bar{\lambda}E)\tilde{x}(t) + e^{-\bar{\lambda}t}Bu(t) + e^{-\bar{\lambda}t}f(t)$. Since $A - \bar{\lambda}E$ has full column rank n, there exists a nonsingular matrix $T \in \mathbb{R}^{l\times l}$ such thtat $T(A - \bar{\lambda}E) = \begin{bmatrix} I_n \\ 0 \end{bmatrix}$, or more precisely

$$\begin{bmatrix} E_1 \\ E_2 \end{bmatrix} \dot{\tilde{x}} = \begin{bmatrix} I_n \\ 0 \end{bmatrix} \tilde{x} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \tilde{u} + \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

with $\begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = TE$, $\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = TB$, $\tilde{u}(t) = e^{-\bar{\lambda}t}u(t)$ and $\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = Tf$ The part (E_1, I_n) is regular since $\operatorname{rank}(\lambda E_1 - I_n) = n$ for $\lambda = 0$ and

$$E_1\dot{\tilde{x}}_1 = \tilde{x} + B_1\tilde{u} + f_1(t)$$

has a unique solution for every sufficiently smooth $\tilde{u}(t)$ and every sufficiently smooth inhomogeneity $f_1(t)$ according to theorem 2.17. From $E_2\dot{\tilde{x}} = B_2\tilde{u} + f_2(t)$ we get a consistency condition for $B_2\tilde{u}$ that must hold for the existence of a solution. Three exists an arbitrary smooth inhomogeneity f_2 for which $E_2\dot{\tilde{x}} \neq B_2\hat{u} + f_2(t)$ and hence the system is not regular.

□

Remark 2.19. Note that for linear descriptor systems with constant coefficient the regularity of (E, A) is advantageous but not necessary (the system can still be consistent). For singular pairs (E, A) we can construct the Kronecker canonical form (KCF) instead of the Weierstraß canonical form (not in this lecture).

2.2 Linear DAEs with variable coefficients

Now we consider descriptor systems of the form

$$\begin{aligned} E(t)\dot{x}(t) &= A(t)x(t) + B(t)u(t) + f(t), \qquad x(t_0) = x_0 \\ y(t) &= C(t)x(t) + D(t)u(t) + g(t). \end{aligned} \tag{2.10}$$

In order to analyze the properties of the system we perform a behavior approach (first suggested by Jan Willems $\sim$ 1990). We introduce the behavior vector $z = \begin{bmatrix} x \\ u \end{bmatrix}$ and write the state equation as

$$\underbrace{\begin{bmatrix} E(t) & 0 \end{bmatrix}}_{:=\mathcal{E}} \dot{z} = \underbrace{\begin{bmatrix} A(t) & B(t) \end{bmatrix}}_{:=\mathcal{A}} z + f(t),$$

or equivalently as

$$\mathcal{E}(t)\dot{z}(t) = \mathcal{A}(t)z(t) + f(t) \tag{2.11}$$

with $\mathcal{E}, \mathcal{A} : \mathbb{I} \to \mathbb{R}^{l\times(n+m)}$.

Remark 2.20. The derivative of the input u occurs only formally in (2.11). Moreover, we could also include the output equation by introducing $z = \begin{bmatrix} x^\top & y^\top & z^\top \end{bmatrix}^\top$ and considering

$$\begin{bmatrix} E(t) & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \dot{z} = \begin{bmatrix} A(t) & B(t) & 0 \\ C(t) & D(t) & -I_p \end{bmatrix} z + \begin{bmatrix} f(t) \\ g(t) \end{bmatrix}.$$

However, since the output equation explicitly determines y, the output equation will not contribute to the analysis and has not to be considered.

For system (2.11) we can apply the theory for nonsquare linear DAEs with variable coefficients based on the *strangeness-index* concept (see also the DAE lecture). At first we construct the *inflated system* (or *derivative array*) obtained by the original DAE (2.11) and all derivatives

$$\left(\frac{d}{dt}\right)^i (\mathcal{E}(t)\dot{z}(t)) = \left(\frac{d}{dt}\right)^i (\mathcal{A}(t)z(t)) + \left(\frac{d}{dt}\right)^i f(t) \qquad i = 1, \ldots, k$$

up to some order k of the form

$$\mathcal{M}_k(t)\dot{v}_k(t) = \mathcal{N}_k(t)v_k(t) + h_k(t),$$

where $\mathcal{M}_k, \mathcal{N}_k : \mathbb{I} \to \mathbb{R}^{(k+1)l \times (n+m)}$ are given by

$$\begin{aligned}
(\mathcal{M}_k)_{i,j} &= \binom{i}{j} \mathcal{E}^{(i-j)} - \binom{i}{j+1} \mathcal{A}^{(i-j-1)}, && i, j = 0, \ldots, k, \\
(\mathcal{N}_k)_{i,j} &= \begin{cases} \mathcal{A}^{(i)} & , i = 0, \ldots, k, j = 0, \\ 0 & \text{otherwise}, \end{cases} \\
(v_k)_j &= z^{(j)}, && j = 0, \ldots, k, \\
(h_k)_j &= f^{(j)}, && j = 0, \ldots, k.
\end{aligned}$$

Example 2.21. *For $k = 2$, we add the first derivative*

$$\mathcal{E}\ddot{z} + \dot{\mathcal{E}}\dot{z} = \dot{\mathcal{A}}z + \mathcal{A}\dot{z} + \dot{f}$$

and the second derivative

$$\mathcal{E}z^{(3)} + 2\dot{\mathcal{E}}\ddot{z} + \ddot{\mathcal{E}}\dot{z} = \ddot{\mathcal{A}}z + 2\dot{\mathcal{A}}\dot{z} + \mathcal{A}\ddot{z} + \ddot{f}$$

of (2.11) *to the DAE* (2.11) *to obtain*

$$\begin{bmatrix} \mathcal{E} & 0 & 0 \\ \dot{\mathcal{E}} - \mathcal{A} & \mathcal{E} & 0 \\ \ddot{\mathcal{E}} - 2\dot{\mathcal{A}} & 2\dot{\mathcal{E}} - \mathcal{A} & \mathcal{E} \end{bmatrix} \begin{bmatrix} \dot{z} \\ \ddot{z} \\ z^{(3)} \end{bmatrix} = \begin{bmatrix} \mathcal{A} & 0 & 0 \\ \dot{\mathcal{A}} & 0 & 0 \\ \ddot{\mathcal{A}} & 0 & 0 \end{bmatrix} \begin{bmatrix} z \\ \dot{z} \\ \ddot{z} \end{bmatrix} + \begin{bmatrix} f \\ \dot{f} \\ \ddot{f} \end{bmatrix}.$$

Hypothesis 1. *There exists integers $\hat{\mu}, \hat{a}, \hat{d}$ and $\hat{v}$ such that the inflated pair $(\mathcal{M}_{\hat{\mu}}, \mathcal{N}_{\hat{\mu}})$ associated with the pair of matrix-valued functions $(\mathcal{E}(t), \mathcal{A}(t))$ has the following properties:*

1. *For all $t \in \mathbb{I}$ we have*
$$rank\left(\mathcal{M}_{\hat{\mu}}(t)\right) = (\hat{\mu}+1)\, l - \hat{a} - \hat{v}$$
such that there exists a smooth matrix-valued function Z of size $(\hat{\mu}+1)l \times (\hat{a}+\hat{v})$ and pointwise maximal rank satisfying
$$Z^\top \mathcal{M}_{\hat{\mu}} = 0.$$

2. *For all $t \in \mathbb{I}$ we have*
$$rank\left(Z^\top \mathcal{N}_{\hat{\mu}} \begin{bmatrix} I_{n+m} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right) = \hat{a}.$$
This implies that without loss of generality Z can be partitioned as $Z = \begin{bmatrix} Z_2 & Z_3 \end{bmatrix}$ with Z_2 of size $(\hat{\mu}+1)l \times \hat{a}$ and Z_3 of size $(\hat{\mu}+1)l \times \hat{v}$ such that
$$\hat{\mathcal{A}}_2 := Z_2^\top \mathcal{N}_{\hat{\mu}} \begin{bmatrix} I_{n+m} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$
has full row rank $\hat{a}$ and $Z_3^\top \mathcal{N}_{\hat{\mu}} \begin{bmatrix} I_{n+m} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$. Furthermore, there exists a smooth matrix-valued function T_2 of size $(n+m) \times (n+m-\hat{a})$ and pointwise maximal rank satisfying
$$\hat{\mathcal{A}}_2 T_2 = 0.$$
Note that $n+m-\hat{a} = \hat{d}+\hat{u}$, where $\hat{u}$ denotes the number of undetermined components.

3. *For all $t \in \mathbb{I}$ we have*
$$rank(\mathcal{E}(t)T_2(t)) = \hat{d}$$
such that there exists a smooth matrix-valued function Z_1 of size $l \times \hat{d}$, where
$$\begin{aligned} \hat{d} &= l - \hat{a} - v_\mu, \qquad \text{with} \\ v_\mu &= l - rank\left(\begin{bmatrix} \mathcal{M}_{\hat{\mu}} & \mathcal{N}_{\hat{\mu}} \end{bmatrix}\right) + rank\left(\begin{bmatrix} \mathcal{M}_{\hat{\mu}-1} & \mathcal{N}_{\hat{\mu}-1} \end{bmatrix}\right) \end{aligned}$$
with the convention $rank\left(\begin{bmatrix} \mathcal{M}_{-1} & \mathcal{N}_{-1} \end{bmatrix}\right) = 0$. Moreover Z_1 has pointwise maximal rank satisfying
$$rank\left(Z_1^\top(t)\mathcal{E}(t)\right) = \hat{d}.$$

Definition 2.22. The smallest possible $\hat{\mu}$ in Hypothesis 1 is called the *strangeness-index* or *s-index* of the behavior system (2.11). A behavior system (2.11) with $\hat{\mu} = 0$ is called *strangeness-free*

If Hypothesis 1 is satisfied for $\hat{\mu}, \hat{a}, \hat{d}$ and $\hat{v}$ (we say that the s-index is well-defined), we can formulate the reduced system

$$\begin{bmatrix} \hat{\mathcal{E}}_1(t) \\ 0 \\ 0 \end{bmatrix} \dot{z}(t) = \begin{bmatrix} \hat{\mathcal{A}}_1(t) \\ \hat{\mathcal{A}}_2(t) \\ 0 \end{bmatrix} z(t) + \begin{bmatrix} \hat{f}_1(t) \\ \hat{f}_2(t) \\ \hat{f}_3(t) \end{bmatrix} \begin{matrix} (\hat{d}) \\ (\hat{a}) \\ (\hat{v}) \end{matrix} \tag{2.12}$$

with $\hat{\mathcal{E}}_1 = Z_1^\top \mathcal{E}, \hat{\mathcal{A}}_1 = Z_1^\top \mathcal{A}_1, \hat{f}_1 = Z_1^\top f, \hat{\mathcal{A}}_2 = Z_2^\top \mathcal{N}_{\hat{\mu}} \begin{bmatrix} I_{n+m} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \hat{f}_2 = Z_2^\top h_{\hat{\mu}}, \hat{f}_3 = Z_3^\top h_{\hat{\mu}}.$

Remark 2.23.

1. In principle, the state vector z can be partitioned in $\begin{bmatrix} z_1^\top & z_2^\top & z_3^\top \end{bmatrix}^\top$ with $z_1 \in \mathbb{R}^{\hat{d}}$ the *differential components*, $z_2 \in \mathbb{R}^{\hat{a}}$ the *algebraic components* and $z_3 \in \mathbb{R}^{\hat{u}}$ the undetermined components ($\hat{u} = n + m - \hat{d} - \hat{a}$, but this would mix up states x and controls u.

2. For a linear DAE without input, i.e.

   $$E(t)\dot{x}(t) = A(t)x(t) + f(t)$$

   the system is called *regular* if it satisfies Hypothesis 1 for $l = n$ (and $m = 0$) and $\hat{\mu}, \hat{a}, \hat{d}, \hat{v}$ such that $n = \hat{d} + \hat{a}$ (i.e. $\hat{v} = 0$ and $\hat{u} = n - \hat{a} - \hat{d} = 0$). It is called regular and *strangeness-free* if it satisfies Hypothesis 1 with $\hat{\mu} = 0$ and $\hat{v} = \hat{u} = 0$. In the following we will say: The descriptor system (2.10) is regular and strangeness-free as a *free system* if it satisfies Hypothesis 1 for $u(t) \equiv 0$ and $\hat{\mu} = 0, m = 0, l = n = \hat{d} + \hat{a}$.

3. For pairs of constant matrices (E, A) the s-index is always well-defined. The condition for regularity of the pencil $\lambda E - A$ can be replaced by the condition $\hat{v} = \hat{u} = 0$ implying that $l = n + m$. We have the relation that $\nu = \mathrm{ind}(E, A) = \hat{\mu} + 1$.

Example 2.24. *Consider the system*

$$\begin{bmatrix} 0 & 0 \\ 1 & -t \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} -1 & t \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

with $n = l = 2$. Hypothesis 1 is not satisfied for $\hat{\mu} = 0$. For $\hat{\mu} = 1$ we consider

$$(\mathcal{M}_1, \mathcal{N}_1) = \left(\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & -t & 0 & 0 \\ 1 & -t & 0 & 0 \\ 0 & -1 & 1 & -t \end{bmatrix}, \begin{bmatrix} -1 & t & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right),$$

which has the following properties

1. *For all* $t \in \mathbb{I}$ *we have* $rank(M_1) = 2$. *This implies* $2l - \hat{a} - \hat{v} = 2 \implies \hat{a} + \hat{v} = 2$. *Choosing* $Z^\top = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix}$ *we get* $Z^\top M_1 = 0$.

2. *We compute*

$$rank\left(Z^\top \mathcal{N}_1 \begin{bmatrix} I_2 \\ 0 \end{bmatrix}\right) = rank\left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} -1 & t \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = rank\left(\begin{bmatrix} -1 & t \\ 0 & -1 \end{bmatrix}\right) = 2 = \hat{a}$$

for all $t \in \mathbb{I}$. *Henceforth, we have* $\hat{v} = 0, \hat{d} = 0$ *and set* $Z_2 = Z$ *and* $\hat{\mathcal{A}}_2 = \begin{bmatrix} -1 & t \\ 0 & -1 \end{bmatrix}$. *Choosing* $[\cdot] = T_2 \in \mathbb{R}^{2\times 0}$ *(the empty matrix) we have* $\hat{\mathcal{A}}_2 T_2 = [\cdot]$.

3. *Finally,* $rank(\mathcal{E}T_2) = rank([\cdot]) = 0 = \hat{d}$. *Analogously we can choose* $Z_1 = [\cdot] \in \mathbb{R}^{2\times 0}$ *and Hypothesis 1 is satisfied for* $\hat{\mu} = 1, \hat{a} = 2, \hat{d} = \hat{v} = 0$. *The corresponding reduced system is given by*

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} -1 & t \\ 0 & -1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} \hat{f}_1 \\ \hat{f}_2 \end{bmatrix},$$

with $\begin{bmatrix} \hat{f}_1 \\ \hat{f}_2 \end{bmatrix} = Z^\top \begin{bmatrix} f_1 & f_2 & \dot{f}_1 & \dot{f}_2 \end{bmatrix}^\top = \begin{bmatrix} f_1 \\ f_2 - \dot{f}_1 \end{bmatrix}$. *Simple computations show that the result is given by*

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} tf_2 - t\dot{f}_1 + f_1 \\ f_2 - \dot{f}_1. \end{bmatrix}$$

Remark 2.25.

1. In the reduced system (2.12) the third block row has $\hat{v}$ equations. Note that $\hat{v}$ is in general larger than v_μ, where
$$l = \hat{d} + \hat{a} + v_\mu.$$

2. The reduced system (2.12) is strangeness-free, i.e., it satisfies Hypothesis 1 for $\hat{\mu} = 0$.

Theorem 2.26 (Existence and Uniqueness)**.** *Let the strangeness-index of* $(\mathcal{E}, \mathcal{A})$ *as in* (2.11) *be well-defined (i.e.* $(\mathcal{E}, \mathcal{A})$ *satisfies Hypothesis 1 with constant values* $\hat{\mu}, \hat{a}, \hat{d}, \hat{v}$*) and let* $f \in \mathcal{C}^{\hat{\mu}+1}(\mathbb{I}, \mathbb{R}^l)$. *Then we have:*

1. *The system* (2.11) *is solvable if and only if* $\hat{f}_3(t) \equiv 0$ *in* (2.12).

2. *An initial condition* $z(t_0) = z_0$ *is consistent with the system if and only if* $\hat{\mathcal{A}}_2(t_0)z_0 + \hat{f}_2(t_0) = 0$.

3. *The corresponding IVP is uniquely solvable if and only if in addition* $\hat{u} = 0$.

Remark 2.27. Note that the behavior system (2.11) has the same solution set as the reduced (strangeness-free) system (2.12) (since the variable z stays the same).

In the original control setting, the reduced formulation (2.12) takes the form

$$\begin{aligned} E_1(t)\dot{x}(t) &= A_1(t)x(t) + B_1(t)u(t) + \hat{f}_1(t) && (\hat{d}) \\ 0 &= A_2(t)x(t) + B_2(t)u(t) + \hat{f}_2(t) && (\hat{a}) \\ 0 &= \hat{f}_3(t) && (\hat{v}) \\ y &= C(t)x(t) + D(t)u(t) + g(t) && (p) \end{aligned} \tag{2.13}$$

with $E_1(t) = \hat{\mathcal{E}}_1 \begin{bmatrix} I_n \\ 0 \end{bmatrix}$, $A_i = \hat{\mathcal{A}}_i \begin{bmatrix} I_n \\ 0 \end{bmatrix}$, $B_i = \hat{\mathcal{A}}_i \begin{bmatrix} 0 \\ I_m \end{bmatrix}$ for $i = 1, 2$.

Remark 2.28.

1. The submatrix $\hat{\mathcal{A}}_2$ has been obtained from the block matrix

$$\begin{bmatrix} A & B \\ \dot{A} & \dot{B} \\ \vdots & \vdots \\ A^{(\hat{\mu})} & B^{(\hat{\mu})} \end{bmatrix}$$

   by transformations from the left only. Therefore, we only need the derivatives of the coefficient matrices, but no derivatives of the input function u (the derivatives of u occur only formally in the inflated pair, but are not used for the construction of (2.13).) Thus, we need no further smoothness requirements for the input u.

2. Since only transformations from the left are used the part stemming from the original states x and the part from the original inputs u is not mixed up.

3. Also initial conditions stay the same.

Lemma 2.29. *A DAE of the form*

$$\begin{bmatrix} \mathcal{E}_1(t) \\ 0 \\ 0 \end{bmatrix} \dot{z}(t) = \begin{bmatrix} \mathcal{A}_1(t) \\ \mathcal{A}_2(t) \\ 0 \end{bmatrix} z(t) = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \qquad \begin{matrix} (d) \\ (a) \\ (v) \end{matrix}$$

is strangeness-free if and only if the matrix

$$\begin{bmatrix} \mathcal{E}_1(t) \\ \mathcal{A}_2(t) \end{bmatrix}$$

has pointwise full row rank $a + d$ for all $t \in \mathbb{I}$.

Proof. The proof is left as exercise. □

Example 2.30. *Consider the system*

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \qquad x(t_0) = x_0.$$

In behavior form, the system is given by

$$(\mathcal{E}, \mathcal{A}) = \left(\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right], \left[\begin{array}{cc|c} 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}\right] \right) = \left(\begin{bmatrix} \mathcal{E}_1 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathcal{A}_1 \\ \mathcal{A}_2 \end{bmatrix} \right).$$

We have $\operatorname{rank}\left(\begin{bmatrix} \mathcal{E}_1 \\ \mathcal{A}_2 \end{bmatrix} \right) = \operatorname{rank}\left(\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \right) = 2$ *and hence the system is strangeness-free in the behavior form (i.e.* $\hat{\mu} = 0$*). However, for a given input* $u(t)$ *(e.g. we consider the free system with* $u(t) \equiv 0$*) we have to consider the DAE with*

$$(E, A) = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right) = \left(\begin{bmatrix} E_1 \\ 0 \end{bmatrix}, \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \right).$$

Here we have $\operatorname{rank}\left(\begin{bmatrix} E_1 \\ A_2 \end{bmatrix} \right) = 1 < 2$ *and the system is not strangeness-free as (free) DAE. In particular,* x_2 *is not uniquely determined and hence the system is not regular.*

Definition 2.31. Two pairs of matrix valued functions $(E(t), A(t))$ and $(\tilde{E}(t), \tilde{A}(t))$ with $E, A, \tilde{E}, \tilde{A} : \mathbb{I} \to \mathbb{R}^{m \times n}$ are called *globally equivalent* if there exist pointwise nonsingular matrix functions $P \in \mathcal{C}(\mathbb{I}, \mathbb{R}^{m \times m})$ and $Q \in \mathcal{C}^1(\mathbb{I}, \mathbb{R}^{n \times n})$ such that

$$(E, A) \sim (\tilde{E}, \tilde{A}) = (PEQ, PAQ - PE\dot{Q}).$$

Remark 2.32. The additional term for the matrix A is based on the following fact. Substitute $x = Q\tilde{x}$ in $E\dot{x} = Ax + f$ to end up with

$$E\dot{Q}x + EQ\dot{x} = AQx + f \iff EQ\dot{x} = (AQ - E\dot{Q})x + f.$$

Theorem 2.33. *Under some constant rank assumptions (see* (A1), (A2) *below), the reduced formulation* (2.13) *of the linear descriptor system* (2.10) *is globally equivalent to a control system of the form*

$$\begin{array}{rlllllll}
\dot{x}_1 = & & A_{13}(t)x_3 + & A_{14}(t)x_4 & + & B_{12}(t)u_2 + & f_1(t) & (\hat{d}) \\
0 = & x_2 & & & + & B_{22}(t)u_2 + & f_2(t) & (\hat{a} - \phi) \\
0 = A_{31}(t)x_1 & & & & + \quad u_1 & & + \; f_3(t) & (\phi) \\
0 = & & & & & & + \; f_4(t) & (\hat{v}) \\
y_1 = & & x_3 & & & D_{12}(t)u_2 + & g_1(t) & (\omega) \\
y_2 = C_{21}(t)x_1 & + C_{22}(t)x_2 & & & & D_{22}(t)u_2 + & g_2(t) & (p - \omega).
\end{array} \tag{2.14}$$

Remark 2.34. In the construction of (2.14) we want to avoid transformations that mix x and u. Thus, we are restricted in the choice of possible equivalence transformations.

Proof. We have to consider the matrix pair of the DAE in (2.13) in $\begin{bmatrix} x^\top & u^\top & y^\top \end{bmatrix}^\top$

$$(\hat{\mathcal{E}}, \hat{\mathcal{A}}) = \left(\begin{bmatrix} E_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} A_1 & B_1 & 0 \\ A_2 & B_2 & 0 \\ 0 & 0 & 0 \\ C & D & -I_p \end{bmatrix} \right) \begin{matrix} \hat{d} \\ \hat{a} \\ \hat{v}, \\ p \end{matrix}$$

where E_1 has pointwise full row rank $\hat{d}$ (due to the construction). Then there exists pointwise orthogonal functions $U \in \mathcal{C}(\mathbb{I}, \mathbb{R}^{m\times m})$ and $V \in \mathcal{C}^1(\mathbb{I}, \mathbb{R}^{\hat{d}\times\hat{d}})$ such that $U^\top E_1 V = \begin{bmatrix} \Sigma_1 & 0 \end{bmatrix}$ with $\Sigma_1 \in \mathcal{C}(\mathbb{I}, \mathbb{R}^{\hat{d}\times\hat{d}})$ pointwise nonsingular. After renaming of matrices (which we do without further emphasis) we arrive at

$$\begin{aligned}(\hat{\mathcal{E}}, \hat{\mathcal{A}}) &\sim \left(\begin{bmatrix} \Sigma_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} A_{11} & A_{12} & B_1 & 0 \\ A_{21} & A_{22} & B_2 & 0 \\ 0 & 0 & 0 & 0 \\ C_1 & C_2 & D & -I_p \end{bmatrix} \right) \\ &\sim \left(\begin{bmatrix} I_{\hat{d}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} A_{11} & A_{12} & B_1 & 0 \\ A_{21} & A_{22} & B_2 & 0 \\ 0 & 0 & 0 & 0 \\ C_1 & C_2 & D & -I_p \end{bmatrix} \right).\end{aligned}$$

To proceed, we need the following rank assumption.

$$\text{Assume that } A_{22} : \mathbb{I} \to \mathbb{R}^{\hat{a}\times(n-\hat{d})}) \text{ has pointwise constant rank } \hat{a} - \phi. \tag{A1}$$

Then we can do a column compression similar as above, i.e. there exist global equivalence transformations such that

$$(\hat{\mathcal{E}}, \hat{\mathcal{A}}) \sim \left(\begin{bmatrix} I_{\hat{d}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} A_{11} & A_{12} & A_{13} & B_1 & 0 \\ A_{21} & I_{\hat{a}-\phi} & 0 & B_2 & 0 \\ A_{31} & 0 & 0 & B_3 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ C_1 & C_2 & C_3 & D & -I_p \end{bmatrix} \right) \begin{matrix} \hat{d} \\ \hat{a} - \phi \\ \phi. \\ \hat{v} \\ p \end{matrix}$$

Since $(\hat{\mathcal{E}}, \hat{\mathcal{A}})$ is strangeness-free by construction, the block matrix $[A_2, B_2]$ of the original pair has pointwise full row rank $\hat{a}$. Therefore, $\tilde{B}_3$ of size $\phi \times m$ has pointwise full row rank ϕ and hence

$$(\hat{\mathcal{E}}, \hat{\mathcal{A}}) \sim \left(\begin{bmatrix} I_{\hat{d}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} A_{11} & A_{12} & A_{13} & B_{11} & B_{12} & 0 \\ A_{21} & I_{\hat{a}-\phi} & 0 & B_{21} & B_{22} & 0 \\ A_{31} & 0 & 0 & I_\phi & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \\ C_1 & C_2 & C_3 & D_1 & D_2 & -I_p \end{bmatrix} \right) \begin{matrix} \hat{d} \\ \hat{a} - \phi \\ \phi. \\ \hat{v} \\ p \end{matrix}$$

To proceed, we need a second rank assumption.

Assume that C_3 of size $p \times (n - \hat{d} - \hat{a} + \phi)$ has pointwise constant rank ω. (A2)

Thus, we obtain

$$(\hat{\mathcal{E}}, \hat{\mathcal{A}}) \sim \left(\begin{bmatrix} I_{\hat{d}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & B_{11} & B_{12} & 0 & 0 \\ A_{21} & I_{\hat{a}-\phi} & 0 & 0 & B_{21} & B_{22} & 0 & 0 \\ A_{31} & 0 & 0 & 0 & I_\phi & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ C_{11} & C_{12} & I_\omega & 0 & D_{11} & D_{12} & -I_\omega & 0 \\ C_{21} & C_{22} & 0 & 0 & D_{21} & D_{22} & 0 & -I_{p-\omega} \end{bmatrix} \right).$$

Finally, the identity blocks can be used for block row and column eliminations. Hereby we restrict the column eliminations to those acting only on columns that belong to the same variable x, y and u.

$$(\hat{\mathcal{E}}, \hat{\mathcal{A}}) \sim \left(\begin{bmatrix} I_{\hat{d}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} A_{11} & 0 & A_{13} & A_{14} & 0 & B_{12} & 0 & 0 \\ 0 & I_{\hat{a}-\phi} & 0 & 0 & 0 & B_{22} & 0 & 0 \\ A_{31} & 0 & 0 & 0 & I_\phi & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_\omega & 0 & 0 & D_{12} & -I_\omega & 0 \\ C_{21} & C_{22} & 0 & 0 & 0 & D_{22} & 0 & -I_{p-\omega} \end{bmatrix} \right).$$

In the last step we perform a global equivalence transformation with $P = I$ and $Q = \operatorname{diag}(Q_1, I, \ldots, I)$, such that

$$(\hat{\mathcal{E}}, \hat{\mathcal{A}}) \sim \left(\begin{bmatrix} Q_1 & 0 & \cdots \\ 0 & 0 & \cdots \\ \vdots & & 0 \end{bmatrix}, \begin{bmatrix} A_{11}Q_1 - \dot{Q}_1 & 0 & A_{13} & \cdots \\ 0 & \cdots & & \\ A_{31}Q_1 & \cdots & & \\ \vdots & & & \end{bmatrix} \right).$$

If we choose Q_1 as the solution of the IVP

$$\dot{Q}_1 = A_{11}Q_1, \qquad Q_1(t_0) = I_{\hat{d}} \quad \text{on } \mathbb{I},$$

the unique solvability of the IVP ensures that Q_1 is pointwise nonsingular and $(\hat{\mathcal{E}}, \hat{\mathcal{A}})$ is globally equivalent to the pair in (2.14). □

Now we are able to characterize consistency and regularity of the descriptor system (2.10).

Corollary 2.35. *Let the strangeness-index $\hat{\mu}$ be well-defined for the system* (2.11) *in behavior form. Furthermore, let the quantities ϕ and ω (as defined in the proof of Theorem 2.33) be constant on $\mathbb{I}$. Then, we have the following:*

1. *The linear descriptor system* (2.10) *is consistent if and only if either $\hat{v} = 0$ or $f_4 \equiv 0$.*

2. *If the system is consistent and if $\phi = 0$, then for a given input u an initial condition is consistent if and only if*

$$x_2(t_0) = -B_2 2(t_0)u_2(t_0) - f_2(t_0)$$

(for the system in form (2.14)*) holds.*

3. *The system* (2.10) *is regular and strangeness-free (as a free system with $u(t) \equiv 0$) if and only if $\hat{v} = \phi = 0$ and $\hat{d} + \hat{a} = n$. (In this case there exists a unique solution for every sufficiently smooth u and f and consistent initial conditions, i.e. the system* (2.10) *is regular!)*

Remark 2.36.

1. If the system (2.10) is consistent and $\hat{v} > 0$ the equations corresponding to the fourth block row of (2.14) describe the redundancies in the system that can simply be omitted.

2. Even for a consistent system (2.10) with consistent initial conditions the solution of the corresponding IVP will in general not be unique (since x_3 and x_4 are not determined).

Proof. The global equivalence transformations leading to (2.14) do not change the solution behavior, since we do not mix up the variables x, u and y (there is a one-to-one correspondence between the solution sets of (2.10) and (2.14) via pointwise nonsingular transformations applied separately to x, u and y). Thus, it is sufficient to consider (2.14).

1. If $\hat{v} = 0$ and $f_4 \neq 0$ there clearly exists no solution (independent of the choice of u). Conversely, if $\hat{v} = 0$ or $f_4 \equiv 0$ we can determine u as follows. Setting $u_2 = 0$ and choosing $x_3 = x_4 = 0$ we get x_2 from $x_2(t) = -f_2(t)$ and x_1 as solution of

$$\dot{x}_1 = f_1(t).$$

With this we can set $u_1 = -A_{31}(t)x_1 - f_3(t)$ and have found a solution for x_1, x_2, x_3 and x_4.

2. Let the system be consistent with $\phi = 0$. Then the system reduces to

$$\begin{aligned} \dot{x}_1 &= A_{13}(t)x_3 + A_{14}(t)x_4 + B_{12}(t)u_2 + f_1 \\ 0 &= x_2 + B_{22}(t)u_2 + f_2. \end{aligned}$$

For every fixed input u_2 this is a strangeness-free DAE (due to Lemma 2.29). The second equation represents the algebraic part. Since x_3 and x_4 are undetermined the solution will not be unique (in general).

3. Assume that $\hat{v} = \phi = 0$ and $\hat{d} + \hat{a} = n$. Then (2.14) reduces to

$$\begin{aligned} \dot{x}_1 &= B_{12}(t)u_2 + f_1 \\ 0 &= x_2 + B_{22}(t)u_2 + f_2. \end{aligned}$$

This system is uniquely solvable for every input u_2 and every inhomogeneity and consistent initial conditions. Moreover, it is strangeness-free for $u_2 = 0$. Conversely let the system be regular and strangeness-free for $u = 0$. We have

$$\begin{aligned} \dot{x}_1 &= A_{13}x_3 + A_{14} + f_1 && (\hat{d}) \\ 0 &= x_2 + f_2 && (\hat{a} - \phi) \\ 0 &= A_{31}x_1 + f_3 && (\phi) \\ 0 &= f_4 && (\hat{v}). \end{aligned}$$

The last equation restricts the inhomogeneity, hence $\hat{v} = 0$. If $\phi > 0$ we have either s-index bigger zero (for $A_{31} \neq 0$) or a consistency condition for f_3 ($A_{31} = 0$) and hence $\phi = 0$. If $\hat{d} + \hat{a} \neq n$ there are free solution components, which contradicts the assumption. Hence $\hat{d} + \hat{a} = n$.

□

Example 2.37. *Consider the control problem*

$$\begin{bmatrix} 0 & 0 \\ 1 & \eta t \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -1 & -\eta t \\ 0 & -(1+\eta) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u, \qquad \eta \in \mathbb{R}.$$

The reduced formulation of the behavior system (2.13) *takes the form (see Ex. I.3)*

$$\begin{bmatrix} 1 & \eta t \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & -(1+\eta) \\ -1 & -\eta t \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u.$$

The transformation into the global equivalent form (2.14) *looks as follows*

$$\left(\hat{\mathcal{E}}, \hat{\mathcal{A}}\right) \equiv \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \right) \qquad \text{by using } Q = \begin{bmatrix} 1 & -\eta t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

with $A_{22} = 0$ *of size* $\hat{a} \times (n - \hat{d}) = 1 \times 1$ *of rank* $0 = \hat{d} - \phi$. *This implies* $\phi = 1$. *The system in form* (2.14) *is given by*

$$\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \right) = \left(\begin{bmatrix} I_{\hat{d}} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & A_{14} & 0 \\ A_{31} & 0 & I_\phi \end{bmatrix} \right)$$

with $\hat{d} = 1, \hat{a} = 1, \phi = 1$. *The system is consistent since* $\hat{v} = 0$, *but not regular and strangeness-free as a free system since* $\phi \neq 0$ *(cp. Ex. I.3).*

2.3 Nonlinear systems

Recall the nonlinear system (1.1) of the form

$$\begin{aligned} F(t, x, \dot{x}, u) &= 0, \\ y - G(t, x, u) &= 0. \end{aligned}$$

As in Section 2.2 we use a behavior approach, i.e. we set $z = \begin{bmatrix} x \\ u \end{bmatrix}$ and consider

$$F(t, z, \dot{z}) = 0 \tag{2.15}$$

instead of (1.1). Since the output equation does not influence the consistency and regularity, we do not consider it in our analysis. Similar as before, we introduce a nonlinear derivative array

$$\mathcal{F}_k\left(t, z, \dot{z}, \ldots, z^{(k+1)}\right) = \begin{bmatrix} F(t, z, \dot{z}) \\ \frac{d}{dt}F(t, z, \dot{z}) \\ \vdots \\ \left(\frac{d}{dt}\right)^k F(t, z, \dot{z}) \end{bmatrix} = 0.$$

Also in the nonlinear case we can formulate a Hypothesis.

Hypothesis 2. *Consider a system of nonlinear DAEs* (2.15) *in behavior form. Then there exist integers* μ, r, a, d *and* v *such that the set*

$$\mathbb{L}_\mu = \left\{ z_\mu \in \mathbb{I} \times \mathbb{R}^{n+m} \times \ldots \times \mathbb{R}^{n+m} \middle| \mathcal{F}_\mu(z_\mu) = 0 \right\}$$

is nonempty and such that for every point $z_\mu^0 = \left(t_0, z_0, \dot{z}_0, \ldots, z_0^{(\mu+1)}\right) \in \mathbb{L}_\mu$, *where* $z_0^{(j)}$ *denotes an algebraic variable, there exists a neighborhood of* z_μ^0 *in which the following properties hold:*

1. *The set* $\mathbb{L}_\mu \subseteq \mathbb{R}^{(\mu+2)(n+m)+1}$ *forms a manifold of dimension* $(\mu+2)(n+m)+1-r$.

2. *We have* $rank\left(\mathcal{F}_{\mu,[z,\dot{z},\ldots,z^{(\mu+1)}]}\right) = r$ *on* $\mathbb{L}_\mu$.

3. *We have*
$$corank\left(\mathcal{F}_{\mu,[z,\dot{z},\ldots,z^{(\mu+1)}]}\right) - corank\left(\mathcal{F}_{\mu-1,[z,\dot{z},\ldots,z^{(\mu)}]}\right) = v$$
on $\mathbb{L}_\mu$, *where* $corank(\mathcal{F}_{-1,z}) = 0$ *by convention.*

4. *We have* $rank\left(\mathcal{F}_{\mu,[\dot{z},\ldots,z^{(\mu+1)}]}\right) = r - a$ *on* $\mathbb{L}_\mu$ *and there exist smooth full rank matrix-valued functions* Z_2 *of size* $(\mu+1)l \times a$ *and* T_2 *of size* $(n+m) \times (n+m-a)$ *defined on* $\mathbb{L}_\mu$ *respectively, that satisfy*
$$\begin{aligned} Z_2^\top \mathcal{F}_{\mu,[\dot{z},\ldots,z^{(\mu+1)}]} &= 0, \\ rank\left(Z_2^\top \mathcal{F}_{\mu,z}\right) &= a, \\ Z_2^\top \mathcal{F}_{\mu,z} T_2 &= 0. \end{aligned}$$

5. *We have* $rank(F_{,\dot{z}}T_2) = d = l - a - v$ *on* $\mathbb{L}_\mu$ *and there exists a smooth full rank matrix valued function* Z_1 *defined on* $\mathbb{L}_\mu$ *such that* $Z_1^\top F_{,\dot{z}} T_2$ *has full rank.*

Remark 2.38.

1. Compared to the linear setting we have $\mathcal{M}_\mu \cong \mathcal{F}_{\mu,[z,\dot{z},\ldots,z^{(\mu+1)}]}$ and $\mathcal{N}_\mu \cong \mathcal{F}_{\mu,z}$.

2. For a matrix $A \in \mathbb{R}^{m\times n}$ the $\operatorname{corank}(A) := m - \operatorname{rank}(A)$ is the codimension of the range of A.

3. A nonempty set $\mathbb{L}_\mu \subseteq \mathbb{R}^{\tilde{n}}$ that is locally diffeomorphic to an open set V in $\mathbb{R}^r$, i.e. the set can locally be parametrized by r scalars, is called a manifold of dimension r. This means that for each $z_0 \in \mathbb{L}_\mu$, z_0 partitioned into $\begin{bmatrix} x_0^\top & y_0^\top \end{bmatrix}^\top$, $x_0 \in \mathbb{R}^r, y_0 \in \mathbb{R}^{\tilde{n}-r}$, there exists a neighborhood $V \subseteq \mathbb{R}^r$ of x_0 and a neighborhood $\tilde{U}$ of $z_0 \in \mathbb{L}_\mu$ such that

$$U := \mathbb{L}_\mu \cap \tilde{U} = \left\{ g(x) \middle| x \in V \right\},$$

   where $g : V \to U$ is a diffeomorphism.

Definition 2.39. The smallest possible μ in Hypothesis 2 is called the strangeness-index of the DAE (2.15).

If F is sufficiently smooth and satisfies Hypothesis 2 with μ, r, a, d, v, then (locally) we can derive a reduced system for the form

$$\begin{aligned} \hat{F}_1(t, z, \dot{z}) &= 0 \qquad && (d \text{ differential equations}) \\ \hat{F}_2(t, z) &= 0 && (a \text{ algebraic equations}) \end{aligned} \tag{2.16}$$

with $\hat{F}_1 : \mathbb{I} \times \mathbb{D}_z \times \mathbb{D}_{\dot{z}} \to \mathbb{R}^d$, $\hat{F}_2 : \mathbb{I} \times \mathbb{D}_z \to \mathbb{R}^a$, where $\hat{F}_1 = Z_1^\top F(t, z, \dot{z})$ and $\hat{F}_2 = Z_2^\top \mathcal{F}_\mu(t, z, \mathcal{H}(\tilde{\omega}_0))$.

Remark 2.40.

1. The reduced system (2.16) can be constructed locally using the implicit function theorem.

2. The reduced system (2.16) is strangeness-free, and every solution of the original system (2.15) also solves (2.16). However, (2.16) may still contain underdetermined components since $m + n \geq d + a$.

3. Let $z_\mu^0 = (t_0, z_0, \dot{z}_0, \ldots, z_0^{(\mu+1)}) \in \mathbb{L}_\mu$ be fixed. By Hypothesis 2, $\mathbb{L}_\mu$ is a manifold of dimension $\tilde{n} = (\mu + 2)(n + m) + 1 - r$ that can locally be parametrized by $\tilde{n}$ parameters. Choosing $\tilde{n}$ parameters ω out of $(t, z, \dot{z}, \ldots, z^{(\mu+1)})$, then there exists a neighborhood $V \subseteq \mathbb{R}^{\tilde{n}}$ of $\tilde{\omega}_0$, as part of z_μ^0 corresponding to ω, and a neighborhood $\tilde{U} \subseteq \mathbb{R}^{(\mu+2)(n+m)+1}$ of z_μ^0 such that

$$U := \mathbb{L}_\mu \cap \tilde{U} = \left\{ g(\tilde{\omega}) \middle| \tilde{\omega} \in V \right\},$$

where $g : V \to U$ is a diffeomorphism. Thus $\mathcal{F}_\mu(z_\mu) = 0$ holds locally if and only if $z_\mu = g(\omega)$ for some $\omega \in U$. In particular, there exists a function $\mathcal{H} : V \to \mathbb{R}^{(\mu+1)(n+m)}$ such that

$$\left(\dot{z}, \ldots, z^{(\mu+1)}\right) = \mathcal{H}(\omega) \qquad \text{for all } \omega \in V.$$

This implies $\mathcal{F}_\mu(t, z, \mathcal{H}(\omega)) = 0$ and in particular $\mathcal{F}_\mu(t, z, \mathcal{H}(\tilde{\omega}_0)) = 0$. Thus,

$$\hat{F}_2(t, z) = Z_2^\top \mathcal{F}_\mu(t, z, \mathcal{H}(\tilde{\omega}_0)),$$
$$\hat{F}_1(t, z, \dot{z}) = Z_1^\top F(t, z, \dot{z})$$

are defined locally in a neighborhood of $z_\mu^0 \in \mathbb{L}_\mu$. For more details see the DAE lecture or [6].

We could split z into $\begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix}$ with $z_1(t) \in \mathbb{R}^d$ differential components, $z_2(t) \in \mathbb{R}^a$ algebraic components and $z_3(t) \in \mathbb{R}^u$ undetermined components ($u = m + n - a - d$). But, this would mean we mix input and state variables (u and x) as components of x. Thus, we proceed as follows. From Hypothesis 2 we get

$$\hat{F}_{2,z} T_2 = Z_2^\top \mathcal{F}_{\mu,z} T_2 = 0$$
$$\operatorname{rank}(T_2) = n + m - a$$
$$\operatorname{rank}(\hat{F}_{1,\dot{z}} T_2) = \operatorname{rank}(Z_1^\top F_{,\dot{z}} T_2) = d.$$

Choosing T_2' such that $\begin{bmatrix} T_2' & T_2 \end{bmatrix}$ is nonsingular, we get

$$\operatorname{rank} \begin{bmatrix} \hat{F}_{1,\dot{z}} \\ \hat{F}_{2,z} \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \hat{F}_{1,\dot{z}} T_2' & \hat{F}_{1,\dot{z}} T_2 \\ \hat{F}_{2,z} T_2' & 0 \end{bmatrix} = \operatorname{rank}(\hat{F}_{1,\dot{z}} T_2) + \operatorname{rank}(\hat{F}_{2,z} T_2') = d + a.$$

Thus, $\begin{bmatrix} \hat{F}_{1,\dot{z}}^\top & \hat{F}_{2,z}^\top \end{bmatrix}^\top$ has pointwise full row rank and hence

$$\begin{bmatrix} \hat{F}_{1,\dot{x}} & 0 \\ \hat{F}_{2,x} & \hat{F}_{2,u} \end{bmatrix}$$

of size $(d + a) \times (n + m)$ has full row rank. Note that fixing a control u will in general not give a strangeness-free regular reduced problem, since $\begin{bmatrix} \hat{F}_{1,\dot{x}}^\top & \hat{F}_{2,x}^\top \end{bmatrix}^\top$ may be singular.

Question Can we choose a control u such that the resulting reduced problem is regular and strangeness-free? *Answer:* yes (see TODO 3.3).

As a consequence, we get the following result.

Theorem 2.41. *Let F in* (1.1) *be sufficiently smooth and satisfies Hypothesis 2 with μ, a, d, v. If $v = 0$ and $n = a + d$ and the reduced problem* (2.16) *satisfies the rank condition*

$$\operatorname{rank} \begin{bmatrix} \hat{F}_{1,\dot{x}} \\ \hat{F}_{2,x} \end{bmatrix} = a + d,$$

then the control problem (1.1) *is regular.*

Example 2.42. *We consider the descriptor system*

$$F(t,x,\dot{x},u) = \begin{bmatrix} \dot{x}_2 \\ \log(x_2) + \sin(u) \end{bmatrix} = 0,$$

with $n = 2, m = 1$. *The corresponding behavior system with* $z = \begin{bmatrix} z_1^\top & z_2^\top & z_3^\top \end{bmatrix}^\top = \begin{bmatrix} x_1^\top & x_2^\top & x_3^\top \end{bmatrix}^\top$ *takes the form*

$$F(t,z,\dot{z}) = \begin{bmatrix} \dot{z}_2 \\ \log(z_2) + \sin(z_3) \end{bmatrix} = 0.$$

We check Hypothesis 2 for $\mu = 0$.

$$\mathbb{L}_0 = \left\{ (t, z_1, z_2, z_3, \dot{z}_1, \dot{z}_2, \dot{z}_3) \middle| \ \dot{z}_2 = 0, z_2 = \exp(-\sin(z_3)) \right\} \subseteq \mathbb{R}^7$$

and $\mathbb{L}_0$ *is a manifold of dimension* $5 = (\mu+2)(n+m)+1-r = 7-r$ *and hence* $r = 2$ *(can be parametrized by* $(t, z_1, z_3, \dot{z}_1, \dot{z}_3)$*).*

$$\mathcal{F}_{0,z} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{z_2} & \cos(z_3) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \exp(\sin(z_3)) & \cos(z_3) \end{bmatrix} \qquad \text{on } \mathbb{L}_0$$
$$\mathcal{F}_{0,\dot{z}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus, $\mathrm{rank}(\mathcal{F}_{0,[z,\dot{z}]}) = 2 = r$, $\mathrm{corank}(\mathcal{F}_{0,[z,\dot{z}]}) = 0 = (2-2) = v$ *and* $\mathrm{rank}(\mathcal{F}_{0,\dot{z}}) = 1 = r - a$. *This implies* $a = 1$. *With* $Z_2^\top = \begin{bmatrix} 0 & 1 \end{bmatrix}$ *we obtain* $Z_2^\top \mathcal{F}_{0,\dot{z}} = 0$ *and*

$$\mathrm{rank}(Z_2^\top \mathcal{F}_{0,z}) = \mathrm{rank}\left(\begin{bmatrix} 0 & \exp(\sin(z_3)) & \cos(z_3) \end{bmatrix} \right) = 1 = a.$$

Choosing

$$T_2 = \begin{bmatrix} 1 & 0 \\ 0 & -\cos(z_3) \\ 0 & \exp(\sin(z_3)) \end{bmatrix}$$

we get $Z_2^\top \mathcal{F}_{0,z} T_2 = 0$ *and finally* $\mathrm{rank}(F_{,\dot{z}} T_2) = \mathrm{rank} \begin{bmatrix} 0 & -\cos(z_3) \\ 0 & 0 \end{bmatrix} = 1 = d$ *if e.g.* $z_3 \in [-1,1]$ *(z_3 corresponds to the control). We conclude that Hypothesis 2 is satisfied for* $z_3 \in [-1,1]$ *with* $\mu = 0, a = 1, d = 1, v = 0$ *and with* $Z_1^\top = \begin{bmatrix} 1 & 0 \end{bmatrix}$ *we obtain* $Z_1^\top F_{,\dot{z}} T_2 = \begin{bmatrix} 0 & -\cos(z_3) \end{bmatrix}$ *of rank* $1 = d$. *With these choices for* Z_1, Z_2, *the reduced problem is the same as the original control problem,*

$$F(t,x,\dot{x},u) = \begin{bmatrix} \dot{x}_2 \\ \log(x_2) + \sin(u) \end{bmatrix} = 0, \qquad u(t) \in [-1,1].$$

For the free system with $u(t) = 0$, *we get* $\hat{F} = \begin{bmatrix} \dot{x}_2 \\ \log(x_2) \end{bmatrix} = 0$ *and this system is not strangeness-free, since*

$$\begin{bmatrix} F_{1,\dot{x}} \\ F_{2,x} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & \frac{1}{x_2} \end{bmatrix}$$

is singular. The free system satisfies Hypothesis 2 for $\mu = 1, a = 1, d = 0, v = 1, r = 3$. *In particular,* x_1 *is undetermined.*

CHAPTER

3

FEEDBACK REGULARIZATION

In the control setting, system properties can be modified using feedback control. E.g. one can use

$u = K(t, x)$	with $K : \mathbb{I} \times \mathbb{D}_x \to \mathbb{R}^m$	(state feedback),
$u = F(t)x + w(t)$	with $F \in \mathcal{C}(\mathbb{R}, \mathbb{R}^{m,n}), w : \mathbb{I} \to \mathbb{R}^m$	(proportional state feedback),
$u = Fx + w(t)$	with $F \in \mathbb{R}^{m,n}, w : \mathbb{I} \to \mathbb{R}^m$	(proportional state feedback),
$u = K(t, y)$	with $K : \mathbb{I} \times \mathbb{R}^p \to \mathbb{R}^m$	(output feedback),
$u = F(t)y + w(t)$	with $F \in \mathcal{C}(\mathbb{R}, \mathbb{R}^{m,p}), w : \mathbb{I} \to \mathbb{R}^m$	(proportional output feedback),
$u = Fy + w(t)$	with $F \in \mathbb{R}^{m,p}, w : \mathbb{I} \to \mathbb{R}^m$	(proportional output feedback),
$u = K(t, \dot{x})$	with $K : \mathbb{I} \times \mathbb{D}_{\dot{x}} \to \mathbb{R}^m$	(state derivative feedback),
$u = K(t, \dot{y})$	with $K : \mathbb{I} \times \mathbb{R}^p \to \mathbb{R}^m$	(output derivative feedback)

and combinations of the above are possible. The situation of such *closed-loop systems* is depicted in Figure 3.1. In particular, it is possible to achieve index reduction and regular-

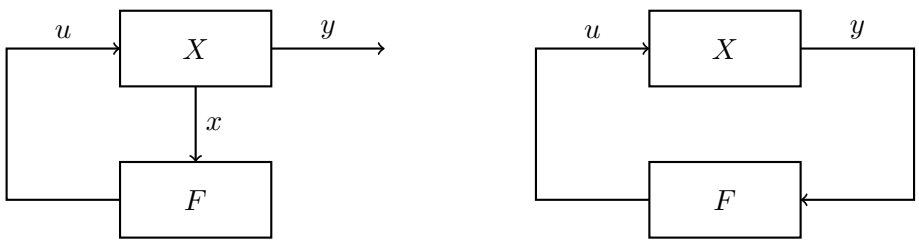


Figure 3.1: Closed-loop systems

ization by using feedback control.

3.1 Linear Descriptor Systems with constant coefficients

Consider system (1.5), given by

$$\begin{aligned} E\dot{x} &= Ax + Bu + f(t), \qquad x(0) = x_0, \\ y &= Cx + Du + g(t). \end{aligned}$$

In the following, we can assume without loss of generality that $D = 0$ in (1.5) (i.e. there is no direct feed-through of the input in the output equation). If $D \neq 0$ we can consider an extended descriptor system of the form

$$\begin{aligned} \begin{bmatrix} E & 0 \\ 0 & 0 \end{bmatrix} \dot{\xi} &= \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} \xi + \begin{bmatrix} B \\ D_2 \end{bmatrix} u, \\ y(t) &= \begin{bmatrix} C & -D_1 \end{bmatrix} \xi, \end{aligned} \tag{3.1}$$

where $D = D_1 D_2$ is a factorization of D (e.g. $D_1 = I, D_2 = D$). The original system (1.5) is equivalent to (3.1) in the sense, that if $x(t)$ is a solution of (1.5) for a given input $u(t)$, then $\xi(t) = \begin{bmatrix} x(t)^\top & -(D_2 u(t))^\top \end{bmatrix}^\top$ solves (3.1). Note that the part $-D_2\dot{u}$ in $\dot{\xi}$ occurs only formally in (3.1). Now, applying a proportional state feedback $u = Fx + w$ to (1.5), we obtain the closed-loop system

$$E\dot{x} = Ax + B(Fx + w) + f = (A + BF)x + Bw + f.$$

Analogously, if we apply proportional output feedback $u = Fy + w$, we obtain the closed-loop system

$$E\dot{x} = Ax + B(Fy + w) + f = Ax + BF(Cx + g) + Bw + f = (A + BFC)x + Bw + BFg + f.$$

Theorem 3.1 (Feedback Regularization)**.** *Consider a control system given by (E, A, B, C).*

1. *There exists a matrix $F \in \mathbb{R}^{m,n}$ such that the matrix pair $(E, A + BF)$ is regular and of index $\nu = ind(E, A + BF) \leq 1$ if and only if $E, A \in \mathbb{R}^{n,n}$ and*

$$rank \begin{bmatrix} E & AS_\infty & B \end{bmatrix} = n,$$

where S_∞ is a matrix with $range(S_\infty) = ker(E)$ (i.e. the columns of S_∞ span the kernel of E).

2. *There exists a matrix $F \in \mathbb{R}^{m,p}$ such that the matrix pair $(E, A + BFC)$ is regular and of index $\nu = ind(E, A + BFC) \leq 1$ if and only $E, A \in \mathbb{R}^{n,n}$ and*

$$rank \begin{bmatrix} E & AS_\infty & B \end{bmatrix} = n \qquad and \qquad rank \begin{bmatrix} E \\ T_\infty^\top A \\ C \end{bmatrix} = n,$$

where S_∞ as above and $range(T_\infty) = ker(E^\top)$.

Example 3.2. *Consider the system*

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$
$$y = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

For the free system without inputs (i.e. $u = 0$) we have to consider

$$(E, A) = \left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right),$$

which is singular. Choose $S_\infty = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$ we have rank $\begin{bmatrix} E & AS_\infty B \end{bmatrix}$ = rank $\begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ = $2 = n$. Thus, choosing $F = \begin{bmatrix} 1 & 0 \end{bmatrix}$, then

$$(E, A + BF) = \left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \right)$$

is regular with $\nu = 1$. With $T_\infty = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top$ we have rank $\begin{bmatrix} E^\top & A^\top T_\infty & C^\top \end{bmatrix}^\top = 1 \neq n$. Thus, there exists no regularizing proportional output feedback.

Proof of Theorem 3.1. The matrices E, A have to be square since otherwise the closed-loop systems cannot be regular. There exist nonsingular matrices $P, Q \in \mathbb{R}^{n,n}$ such that

$$PEQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}, \qquad Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}, \qquad P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$$

and the columns of Q_2 span $\ker(E)$ and the columns of $P_2^\top$ span $\ker(E^\top)$. Setting

$$PAQ = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad PB = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \qquad CQ = \begin{bmatrix} C_1 & C_2 \end{bmatrix},$$

we have

$$\begin{aligned} \operatorname{rank} \begin{bmatrix} E & AQ_2 & B \end{bmatrix} &= \operatorname{rank} \begin{bmatrix} PE & PAQ_2 & B \end{bmatrix} \begin{bmatrix} Q & & \\ & I & \\ & & I \end{bmatrix} \\ &= \operatorname{rank} \begin{bmatrix} PEQ & PAQ_2 & B \end{bmatrix} \\ &= \operatorname{rank} \begin{bmatrix} I_r & 0 & A_{11} & B_1 \\ 0 & 0 & A_{22} & B_2 \end{bmatrix} \\ &= r + \operatorname{rank} \begin{bmatrix} A_{22} & B_2 \end{bmatrix} \overset{!}{=} n. \end{aligned}$$

Thus, the matrix $\begin{bmatrix} A_{22} & B_2 \end{bmatrix}$ must satisfy

$$\operatorname{rank} \begin{bmatrix} A_{22} & B_2 \end{bmatrix} = n - r. \tag{3.2}$$

Analogously, we deduce from

$$\operatorname{rank}\begin{bmatrix} E \\ P_2A \\ C \end{bmatrix} = \operatorname{rank}\begin{bmatrix} PEW \\ P_2AQ \\ CQ \end{bmatrix} = \operatorname{rank}\begin{bmatrix} I_r & 0 \\ 0 & 0 \\ A_{21} & A_{22} \\ C_1 & C_2 \end{bmatrix} = r + \operatorname{rank}\begin{bmatrix} A_{22} \\ C_2 \end{bmatrix} \overset{!}{=} n,$$

that the matrix $\begin{bmatrix} A_{22}^\top & C_2^\top \end{bmatrix}^\top$ must satisfy

$$\operatorname{rank}\begin{bmatrix} A_{22} \\ C_2 \end{bmatrix} = n - r. \tag{3.3}$$

We only need to prove 2 (since 1 follows from $C = I$). The matrix pair $(E, A + BFC)$ is regular and of index ≤ 1 if and only if the matrices are square and $A_{22} + B_2FC_2$ is either not present or nonsingular (see Ex. 1.2). There exist nonsingular matrices $U, V \in \mathbb{R}^{n-r,n-r}$ such that

$$U^{-1}A_{22}V = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}.$$

We set $\bar{A}_{22} := V \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} U^{-1}$. Then it holds that

$$A_{22}\bar{A}_{22}A_{22} = \left(U \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} V^{-1}\right)\left(V \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} U^{-1}\right)\left(U \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} V^{-1}\right) = U \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} V^{-1} = A_{22},$$

and

$$A_{22} + B_2FC_2 = \begin{bmatrix} A_{22} & B_2 \end{bmatrix} \begin{bmatrix} \bar{A}_{22} & 0 \\ 0 & F \end{bmatrix} \begin{bmatrix} A_{22} \\ C_2 \end{bmatrix}. \tag{3.4}$$

Assume that $F \in \mathbb{R}^{n,p}$ exists such that $A_{22} + B_2FC_2$ is nonsingular, then from (3.4) follows (3.2) and (3.3). Conversely, assume that (3.2) and (3.3) hold. Then

$$\operatorname{rank}\left(\begin{bmatrix} A_{22} & B_2FC_2 \end{bmatrix}\right) = \operatorname{rank}\left(U^{-1}\begin{bmatrix} A_{22} & B_2FC_2 \end{bmatrix} V\right) = \operatorname{rank}\left(\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix} F \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 \end{bmatrix}\right)$$
$$= \operatorname{rank}\left(\begin{bmatrix} I & 0 & \tilde{B}_1 \\ 0 & 0 & \tilde{B}_2 \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & F \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & 0 \\ \tilde{C}_1 & \tilde{C}_2 \end{bmatrix}\right)$$

with $U^{-1}B_2 = \begin{bmatrix} \tilde{B}_1^\top & \tilde{B}_2^\top \end{bmatrix}^\top$ and $C_2V = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 \end{bmatrix}$. Note that due to (3.2) and (3.3) the left and right matrix in the last equality have full rank and so have $\tilde{B}_2$ and $\tilde{C}_2$. Hence, there exist nonsingular matrices $\tilde{U}_1, \tilde{U}_2, \tilde{V}_1, \tilde{V}_2$ such that

$$\tilde{U}_1^{-1}\tilde{B}_2\tilde{V}_1 = \begin{bmatrix} I & 0 \end{bmatrix} \qquad \text{and} \qquad \tilde{U}_2^{-1}\tilde{C}_2\tilde{V}_2 = \begin{bmatrix} I \\ 0 \end{bmatrix}.$$

Then

$$\operatorname{rank}\left(A_{22} + B_2FC_2\right) = \operatorname{rank}\left(\begin{bmatrix} I & 0 & B_{11} & B_{12} \\ 0 & 0 & I & 0 \end{bmatrix} \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & F_{11} & F_{12} \\ 0 & 0 & F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & 0 \\ C_{11} & I \\ C_{12} & 0 \end{bmatrix}\right).$$

Choosing $F_{11} = I, F_{12} = 0, F_{21} = 0$ and $F_{22} = 0$ gives

$$\operatorname{rank}(A_{22} + B_2 F C_2) = \operatorname{rank}\left(\begin{bmatrix} I + B_{11}C_{11} & B_{11} \\ C_{11} & I \end{bmatrix}\right) = n - r.$$

□

Remark 3.3. The properties of a control system (1.5) can also be altered by use of derivative feedback control of the form $u = -G\dot{y} + v(t)$ with $G \in \mathbb{R}^{m,p}$ or a combined proportional and derivative feedback $u = Fy - G\dot{y} + v(t)$ leading to the closed-loop system

$$(E + BGC)\dot{x} = (A + BFC)x + Bv + \tilde{f}.$$

For $C = I$ we have the special case of state feedback. The *regularization by feedback* problem then takes the form: For given E, A, B, C find F, G such that $(E+BGC, A+BFC)$ is regular and of index $\nu \leq 1$. For this problem, similar conditions as in the previous theorem can be derived (see e.g. Bunse-Gerstner, Mehrmann, Nichols, 1992).

3.2 Linear Descriptor Systems with variable coefficients

Consider the descriptor system (1.4) given by

$$\begin{aligned} E(t)\dot{x} &= A(t)x + B(t)u + f(t), \qquad x(t_0) = x_0 \\ y &= C(t)x + D(t)u + g(t). \end{aligned}$$

Again, we assume that $D(t) \equiv 0$. In behavior form (2.11) we have

$$\mathcal{E}\dot{z} = \mathcal{A}z + f(t)$$

with $z = \begin{bmatrix} x^\top & u^\top \end{bmatrix}^\top, \mathcal{E} = \begin{bmatrix} E & 0 \end{bmatrix}, \mathcal{A} = \begin{bmatrix} A & B \end{bmatrix}$. If the strangeness index $\hat{\mu}$ of the system in behavior form is well-defined we can formulate the reduced system (2.12) (using Hypothesis 1

$$\begin{bmatrix} \hat{\mathcal{E}}_1(t) \\ 0 \\ 0 \end{bmatrix} \dot{z} = \begin{bmatrix} \hat{\mathcal{A}}_1(t) \\ \hat{\mathcal{A}}_2(t) \\ 0 \end{bmatrix} z + \begin{bmatrix} \hat{f}_1 \\ \hat{f}_2 \\ \hat{f}_3 \end{bmatrix} \qquad \begin{matrix} \hat{d} \\ \hat{a} \\ \hat{v} \end{matrix}$$

with characteristic values $\hat{d}, \hat{a}, \hat{v}$. These characteristic values are invariant under proportional feedback.

Theorem 3.4. *Consider a linear descriptor system of form* (1.4) *and assume that the strangeness index of the system in behavior form* (2.11) *is well-defined. Then, the characteristic values $\hat{d}, \hat{a}$ and $\hat{v}$ are invariant under proportional state feedback $u = F(t)x + w$ and proportional output feedback $\tilde{F}(t)y + w$.*

Proof. In the behavior setting, proportional state feedback takes the form

$$\begin{bmatrix} x \\ u \end{bmatrix} = \underbrace{\begin{bmatrix} I_n & 0 \\ F(t) & I_m \end{bmatrix}}_{=:Q} \begin{bmatrix} \tilde{x} \\ \tilde{u} \end{bmatrix},$$

i.e., is is equivalent to a change of basis of z. Note that this works only, since $\mathcal{E} = \begin{bmatrix} \tilde{E} & 0 \end{bmatrix}$. Similarly, proportional output feedback is an equivalence transformation in the more general behavior approach, where we also include y

$$\begin{bmatrix} x \\ u \\ y \end{bmatrix} = \begin{bmatrix} I_n & 0 & 0 \\ 0 & I_m & F(y) \\ 0 & 0 & I_p \end{bmatrix}$$

and premultiply by $P = \begin{bmatrix} I & B(t)F(t) \\ 0 & I \end{bmatrix}$. Since (global) equivalence transformations of $(\mathcal{E}, \mathcal{A})$ do not change the characteristic quantities, $\hat{\mu}, \hat{a}, \hat{d}$ and $\hat{v}$ are invariant for both types of feedback. □

Corollary 3.5. *Let the strangeness index $\hat{\mu}$ be well-defined for the system* (2.11) *in behavior form and let the quantities ϕ and ω (as in Theorem 2.33) be constant on $\mathbb{I}$. Then, there exists a state feedback $u = F(t)x + w$ such that the closed-loop system*

$$E(t)\dot{x} = (A(t) + B(t)F(t))\, x(t) + B(t)w + f(t) \tag{3.5}$$

is regular (as a free system, i.e. $w \equiv 0$) if and only if $\hat{v} = 0$ and $\hat{d} + \hat{a} = n$.

Proof. Since proportional state feedback can be written as (global) equivalence transformation of the pair $(\mathcal{E}, \mathcal{A})$ in behavior form, it holds that first applying the feedback to the original system (1.4) and then computing the reduced system formulation is the same as first computing the reduced system and then applying the feedback. Thus, the closed-loop system (3.5) is regular as a free system if and only if the reduced formulation (2.12) with inserted feedback is regular and strangeness-free as a free system. Under the assumptions of Theorem 2.33, it is sufficient to consider the system in the form

$$\begin{aligned} \dot{x}_1 &= A_{13}x_3 + A_{14}x_4 + B_{12}u_2 + f_1, && (\hat{d}) \\ 0 &= x_2 + B_{22}u_2 + f_2, && (\hat{a} - \phi) \\ 0 &= A_{31}x_1 + u_1 + f_3, && (\phi) \\ 0 &= f_4. && (\hat{v}) \end{aligned}$$

The output equation is not involved in our considerations, thus we can formally set $\omega = 0$, such that x_3 does not appear. First, we assume that $\hat{v} = 0$ and $\hat{d} + \hat{a} = n$, such that we have to consider

$$\begin{aligned} \dot{x}_1 &= A_{14}x_4 + B_{12}u_2 + f_1, && (\hat{d}) \\ 0 &= x_2 + B_{22}u_2 + f_2, && (\hat{a} - \phi) \\ 0 &= A_{31}x_1 + u_1 + f_3, && (\phi) \end{aligned} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ x_4 \end{bmatrix} \begin{matrix} \hat{d} \\ \hat{a} - \phi. \\ \phi \end{matrix}$$

Thus, x_4 and u_1 are both of size ϕ. Applying the proportional state feedback

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} x_4 - A_{31}x_1 + w_1 \\ w + 2 \end{bmatrix},$$

we get the closed-loop system

$$\begin{aligned} \dot{x}_1 &= A_{14}x_4 + B_{12}w_2 + f_1, && (\hat{d}) \\ 0 &= x_2 + B_{22}w_2 + f_2, && (\hat{a} - \phi) \\ 0 &= x_4 + w_1 + f_3, && (\phi) \end{aligned}$$

and for $w = \begin{bmatrix} w_1^\top & w_2^\top \end{bmatrix}^\top = 0$ we have

$$\begin{aligned} \dot{x}_1 &= A_{14}x_4 + f_1, && (\hat{d}) \\ 0 &= x_2 + f_2, && (\hat{a} - \phi) \\ 0 &= x_4 + f_3, && (\phi) \end{aligned}$$

which is regular and strangeness-free. Conversely, let (3.5) be regular as a free system with $w = 0$. Then, necessarily, we need $\hat{v} = 0$ and $l = n = \hat{a} + \hat{d}$. □

Corollary 3.6. *Let the strangeness index $\hat{\mu}$ be well-defined for the system* (2.11) *in behavior form and let the quantities ϕ and ω (as in Theorem 2.33) be constant on $\mathbb{I}$. There exists an output feedback $u = F(t)y + w$ such that the closed-loop system*

$$E(t)\dot{x} = \left(A(t) + B(t)F(t)C(t)\right)x + B(t)w + f(t) + B(t)F(t)g(t)$$

is regular (as a free system) if and only if $\hat{v} = 0$ and $\hat{d} + \hat{a} = n$ and $\phi = \omega$.

Proof. As in the proof of Corollary 3.5 we can consider the reduced formulation (2.14) for $D = 0$ instead of (1.4).

$\Longleftarrow$: Assume that $\hat{v} = 0, \hat{d} + \hat{a} = n$ and $\phi = \omega$ in (2.14). Thus, the unknown x_4 does not appear in (2.14) and u_1 and y_1 are of the same size. Using the feedback

$$\begin{aligned} u_1 &= y_1 + w_1, \\ u_2 &= w_2, \end{aligned}$$

yields the closed-loop system

$$\begin{aligned} \dot{x}_1 &= A_{13}x_3 + B_{12}w_2 + f_1, && (\hat{d}) \\ 0 &= x_2 + B_{22}w_2 + f_2, && (\hat{a} - \phi) \\ 0 &= A_{31}x_1 + y_1 + w_1 + f_3 = A_{31}x_1 + x_3 + w_1 + f_3 + g_1. && (\phi). \end{aligned}$$

For $w = \begin{bmatrix} w_1^\top & w_2^\top \end{bmatrix}^\top = 0$ this system is regular and strangeness-free, since

$$\operatorname{rank}\left(\begin{bmatrix} E_1 \\ A_2 \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} I_{\hat{d}} & 0 & 0 \\ 0 & I_{\hat{a}-\phi} & 0 \\ A_{31} & 0 & I_\phi \end{bmatrix}\right) = \hat{d} + \hat{a} = n.$$

$\Longrightarrow$: Let the closed-loop system be regular as a free system. Then, necessarily, $\hat{v} = 0$ and $\hat{a} + \hat{d} = n$. For a feedback matrix

$$F = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \quad \begin{matrix} \phi \\ m - \phi \end{matrix},$$

we obtain $A + BFC$ as

$$\begin{aligned} A + BFC &= \begin{bmatrix} 0 & 0 & A_{13} & A_{14} \\ 0 & I_{\hat{a}-\phi} & 0 & 0 \\ A_{31} & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & B_{12} \\ 0 & B_{22} \\ I_\phi & 0 \\ I_\phi & 0 \end{bmatrix} \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} 0 & 0 & I_\omega & 0 \\ C_{21} & C_{22} & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} B_{21}F_{22}C_{21} & B_{12}F_{22}C_{22} & A_{13} + B_{12}F_{12} & A_{14} \\ B_{22}F_{22}C_{21} & I + B_{22}F_{22}C_{22} & B_{22}F_{21} & 0 \\ A_{31} + F_{12}C_{21} & F_{12}C_{22} & F_{11} & 0 \end{bmatrix} \begin{matrix} (\hat{d}) \\ (\hat{a} - \phi) \\ (\phi) \end{matrix} \end{aligned}$$

of size $(\hat{d} + \hat{a}) \times (\hat{d} + \hat{a})$. Thus, the closed-loop system is regular and strangeness-free for $w = 0$ if and only if

$$\begin{bmatrix} I_{\hat{d}} & 0 & 0 & 0 \\ B_{22}F_{22}C_{21} & I + B_{22}F_{22}C_{22} & B_{22}F_{21} & 0 \\ A_{31} + F_{12}C_{21} & F_{12}C_{22} & F_{11} & 0 \end{bmatrix}$$

has pointwise full row rank (is piecewise nonsingular). This implies $\omega - \phi = 0 \iff \phi = \omega$.

□

3.3 Nonlinear Descriptor Systems

Consider system (1.1) given by

$$\begin{aligned} F(t, x, \dot{x}, u) &= 0, \qquad x(t_0) = x_0 \\ y - G(t, x) &= 0 \end{aligned}$$

and the corresponding reduced system formulation (2.16) given by

$$\begin{aligned} \hat{F}_1(t, x, \dot{x}, u) &= 0 \qquad x(t_0) = x_0 \\ \hat{F}_2(t, x, u) &= 0 \\ y - G(t, x) &= 0 \end{aligned}$$

obtained by applying Hypothesis 2. Now the question is: Can we find a feedback control u such that the reduced problem is regular and strangeness-free? Necessarily, we need $n = d + a$. Applying a state feedback $u = K(t, x)$ leads to the closed-loop system

$$\begin{aligned} \hat{F}_1(t, x, \dot{x}, K(t, x)) &= 0 \\ \hat{F}_2(t, x, K(t, x)) &= 0. \end{aligned} \tag{3.6}$$

This closed-loop system is regular and strangeness-free if and only if

$$\begin{bmatrix} \hat{F}_{1,\dot{x}} \\ \hat{F}_{2,x} + \hat{F}_{2,u}K_{,x} \end{bmatrix} \qquad \text{is pointwise nonsingular.} \tag{3.7}$$

Since the reduced system (2.16) is defined only locally, it is sufficient to satisfy condition (3.7) only locally. Thus, we can restrict to linear state feedback $\tilde{K}x(t) + w(t)$ such that $\tilde{K} = K_{,x}$. In Section 2.3 we have seen that

$$\begin{bmatrix} \hat{F}_{1,\dot{x}} & 0 \\ \hat{F}_{2,x} & \hat{F}_{2,u} \end{bmatrix}$$

has full row rank $d + a$. Thus using $E_1(t) := \hat{F}_{1,\dot{x}}, A_2(t) := \hat{F}_{2,x}, B_2(t) := \hat{F}_{2,u}$ similar as in (2.13), the existence of a suitable feedback matrix $\tilde{K} = K_{,x}$ follows from Corollary 3.5. The control function $w(t)$ can be used to satisfy initial conditions of the form

$$u^{(\ell)}(t_0) = \tilde{K}x_0^{(\ell)} + w^{(\ell)}(t_0) = u_0^{(\ell)} \qquad \text{for } \ell = 0, \ldots, \mu + 1.$$

Altogether, we have proved the following theorem.

Theorem 3.7. *Assume that the control problem* (1.1) *in behavior form satisfies Hypothesis 2 with characteristic values values μ, a, d, v and let $d + a = n$. Furthermore, let*

$$z_{\mu,0} = (t_0, x_0, u_0, \ldots, x_0^{(\mu+1)}, u_0^{(\mu+1)}) \in \mathbb{L}_\mu.$$

Then there (locally) exists a state feedback $u = K(t, x)$ satisfying

$$\begin{aligned} u_0 &= K(t_0, x_0) \qquad \textit{and} \\ \dot{u}_0 &= K_{,t}(t_0, x_0) + K_{,x}(t_0, x_0)\dot{x}_0, \end{aligned}$$

such that the closed-loop reduced problem (3.6) *is regular and strangeness-free.*

Example 3.8. *Consider the descriptor system from Example 2.42 given by*

$$F(t, x, \dot{x}, u) = \begin{bmatrix} \dot{x}_2 \\ \log(x_2) + \sin(u) \end{bmatrix} = 0$$

already in reduced form. We have already see in Example 2.42 that the free system with $u(t) \equiv 0$ is not strangeness-free. To obtain a regular and strangeness-free closed-loop system we need to find $\tilde{K}$ such that

$$\begin{bmatrix} \hat{F}_{1,\dot{x}} \\ \hat{F}_{2,x} + \hat{F}_{2,u}K_{,x} \end{bmatrix}$$

is nonsingular. We have

$$\begin{bmatrix} \hat{F}_{1,\dot{x}} & 0 \\ \hat{F}_{2,x} & \hat{F}_{2,u} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{x_2} & \cos(u) \end{bmatrix}.$$

At $z_{0,0} = (t_0, x_{1,0}, x_{2,0}, u_0, \dot{x}_{1,0}, \dot{x}_{2,0}, \dot{u}_0)^\top \in \mathbb{L}_0$ *given by* $z_{0,0} = (0,0,1,0,0,0,0)$ *we have*

$$\begin{bmatrix} \hat{F}_{1,\dot{x}} & 0 \\ \hat{F}_{2,x} & \hat{F}_{2,u} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

and choosing $\tilde{K} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ *gives*

$$\begin{bmatrix} \hat{F}_{1,\dot{x}} \\ \hat{F}_{2,x} + \hat{F}_{2,u} K_{,x} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

nonsingular at $z_{0,0}$*, i.e.* $u(t) = \tilde{K}x + w(t) = x_1 + w(t)$*. Observe that*

$$\begin{aligned} 0 = u_0 = u(t_0) = x_1(t_0) + w(t_0) = x_{1,0} + w(t_0) = w(t_0) \\ 0 = \dot{u}_0 = \dot{u}(t_0) = \dot{x}_1(t_0) + \dot{w}(t_0) = \dot{x}_{1,0} + \dot{w}(t_0) = \dot{w}(t_0). \end{aligned}$$

Thus, we can choose $w(t) \equiv 0$*. The corresponding closed-loop system is given by*

$$\begin{aligned} \dot{x}_2 &= 0, \\ 0 &= \log(x_2) + \sin(x_2), \end{aligned}$$

which is regular and strangeness-free in a neighborhood of $z_{0,0}$*. For* $x_1(0) = x_{1,0} = 0, x_2(0) = x_{2,0} = 1$ *we get the unique solution*

$$\begin{aligned} x_1(t) &\equiv 0 \\ x_2(t) &\equiv 1. \end{aligned}$$

Now, we want to consider also output feedback of the form $u = K(t, y)$. Inserting this control in the reduced problem (2.16) yields the closed-loop reduced problem

$$\begin{aligned} \hat{F}_1(t, x, \dot{x}, K(t, G(t, x))) &= 0 \\ \hat{F}_2(t, x, K(t, G(t, x))) &= 0. \end{aligned}$$

Again, the condition

$$\begin{bmatrix} \hat{F}_{1,\dot{x}} \\ \hat{F}_{2,x} + \hat{F}_{2,u} K_{,y} G_{,x} \end{bmatrix} = \begin{bmatrix} \hat{F}_{1,\dot{x}} & 0 \\ \hat{F}_{2,x} & \hat{F}_{2,u} \end{bmatrix} \begin{bmatrix} I \\ K_{,y} G_{,x} \end{bmatrix} \qquad \text{is nonsingular} \tag{3.8}$$

has to be satisfied. For $y = x$ we are again in the state feedback case. Again, we can proceed as in Section 3.2 for linear systems. Setting $E_1 := \hat{F}_{1,\dot{x}}, A_2 := \hat{F}_{2,x}, B_2 := \hat{F}_{2,u}$ and $C := G_{,x}$. We can determine the quantities ϕ and ω locally at a point $(t_0, z_0, \dot{z}_0)$ given $z_{\mu,0} \in \mathbb{L}_\mu$ as in Theorem 2.33. In this way we get a nonlinear version of Corollary 3.6.

Corollary 3.9. *Suppose that the control problem* (1.1) *in generalized behavior form using*

$$z = \begin{bmatrix} x^\top & u^\top & y^\top \end{bmatrix}^\top$$

satisfies Hypothesis 2 with characteristic values μ, a, d, v *and* $d + a = n$*. Furthermore, let* $\phi = \omega$ *locally for the system at* $(t_0, z_0, \dot{z}_0)$ *given by* $z_{\mu,0} \in \mathbb{L}_\mu$*. Then there exists an output feedback* $u = K(t, y)$ *satisfying* $u_0 = K(t_0, y_0)$ *and* $\dot{u}_0 = K_{,t}(t_0, y_0) + K_{,y}(t_0, y_0)\dot{y}_0$ *such that the closed-loop reduced problem is regular and strangeness-free.*

Proof. Under the given assumptions we can apply the same theory as for the linear problem to obtain a suitable matrix $\tilde{K}_{,y}$ sucht that (3.8) holds. Then, we can use the linear output feedback $u(t) = \tilde{K}y(t) + w(t)$, where the control function $w(t)$ is used to satisfy the given initial conditions. □

CHAPTER

4

CONTROL THEORETICAL CONCEPTS

4.1 Controllability

In contrast to standard state-space systems there are several different notions of controllability for descriptor systems. Unfortunately, there is no uniform terminology in the literature. We consider system (1.1) of the form

$$\begin{aligned} F(t, x, \dot{x}, u) &= 0 \qquad x(t_0) = x_0, \\ y - G(t, x) &= 0. \end{aligned}$$

Definition 4.1.

1. The descriptor system (1.1) is called *completely controllable (C-controllable)* if for any given initial state $x_0 \in \mathbb{R}^n$ and final state $x_f \in \mathbb{R}^n$ there exists a control input u that transforms the system from x_0 to x_f in finite time $t_f \geq t_0$ (i.e. $\exists u, t_f < \infty$ such that $x(t_f; u, x_0) = x_f$).

2. For (1.1) a set $\mathcal{R}_{x_0}$ is called *reachable from* $x_0 \in \mathbb{R}^n$ if for all $x_f \in \mathcal{R}_{x_0}$ there exists an admissible control input u that transfers the system from x_0 to x_f in finite time (i.e. $\exists u, t_f < \infty$ such that $x(t_f; u, x_0) = x_f \in \mathcal{R}_{x_0}$).

$$\mathcal{R}_{x_0} := \left\{ x_f \in \mathbb{R}^n \middle| \; \exists \; u, t_f < \infty : x(t_f : u, x_0) = x_f \right\} \subseteq \mathbb{R}^n.$$

Let $\mathcal{R} := \bigcup_{x_0 \in \mathcal{X}_c^{t_0}} \mathcal{R}_{x_0}$ denote the *reachable set* (where $\mathcal{X}_c^{t_0} \subseteq \mathbb{R}^n$ is the set of all consistent initial values x_0 at time t_0). The system (1.1) is called *controllable within the reachable set (R-controllable)*, if any state in $\mathcal{R}$ can be reached from any consistent initial state x_0 in finite time (i.e. for any $x_0 \in \mathcal{R}, x_f \in \mathcal{R}\ \exists\ u, t_f < \infty$ such that $x(t_f; u, x_0) = t_f$).

Remark 4.2.

1. R-controllability is sometimes also called *finite dynamics controllability.*

2. In general, descriptor systems will not be C-controllable, since algebraic constraints fix the solution onto a certain solution manifold. In the case $E = I_n$, R-controllability coincides with C-controllability.

Example 4.3. *Consider the descriptor system*

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u.$$

The reachable set is given by $\mathcal{R} = \{(x_1, x_2) \in \mathbb{R}^2 \big|\ x_2 = 0\}$ *and the system is R-controllable.*

For descriptor systems another phenomenon arises if input functions are used that are only piecewise continuous. Since the solution may depend on derivatives of the input it may happen that no classical solution exists. Thus, a descriptor system can adopt a generalized solution (i.e. a distribution as solution).

Example 4.4. *Consider the descriptor system*

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix} u.$$

Let the input u be given by

$$u(t) = \begin{cases} 0, & 0 \le t \le 1, \\ 1, & 1 \le t \le t_f, \end{cases}$$

i.e. u is only piecewise continuous. The solution is given by

$$\begin{aligned} x_1(t) &= u + \dot{u} \\ x_2(t) &= u. \end{aligned}$$

Thus, for the given u, no solution in the classical sense exists. Nevertheless, the state response of the system can be depicted as in Figure 4.1. and (x_1, x_2) is a solution in the distributional sense.

Definition 4.5 (first rough version)**.** A descriptor system (1.1) is called *impulse controllable (I-controllable)* if for any given initial state $x_0 \in \mathbb{R}^n$ there exists an admissible control input u that transforms the system to some impulsive state in finite time.

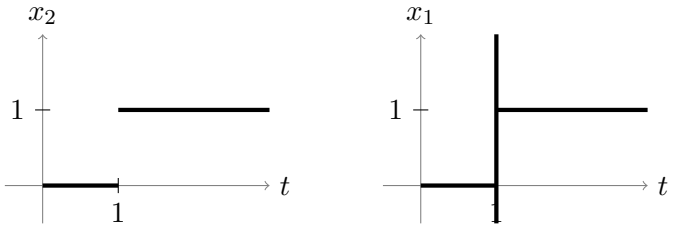


Figure 4.1: Impulse behavior of solutions. The impuse in x_1 is due to the jump behavior in the input.

It can be shown that I-controllability is equivalent to the ability to cancel all impulsive states by choosing a suitable u. This can be done by state feedback control (i.e. for every initial state x_0 there exists a state feedback control such that the closed-loop system has no impulsive solutions).

For regular time-invariant (LTI) descriptor systems of the form

$$\begin{aligned} E\dot{x} &= Ax + Bu, \qquad x(0) = x_0 \\ y &= Cx \end{aligned} \tag{4.1}$$

with $E, A \in \mathbb{R}^{n,n}, B \in \mathbb{R}^{n,m}, C \in \mathbb{R}^{p,n}$, there exists purely algebraic characterizations of the different controllability concepts. Without loss of generality, we assume that $r = \operatorname{rank}(E) < n$. Then there exist nonsingular matrices $T, W \in \mathbb{R}^{n,n}$ such that the system is (strongly) equivalent to

$$\dot{x}_1 = Jx_1 + B_1u, \qquad x_1(0) = x_{1,0} \tag{4.2a}$$

$$N\dot{x}_2 = x_2 + B_2u, \qquad x_2(0) = x_{2,0} \tag{4.2b}$$

$$y = C_1x_1 + C_2x_2. \tag{4.2c}$$

with

$$WET = \begin{bmatrix} I_{n_f} & 0 \\ 0 & N \end{bmatrix}, \qquad WAT = \begin{bmatrix} J & 0 \\ 0 & I_{n_\infty} \end{bmatrix},$$

$$CT = \begin{bmatrix} C_1 & C_2 \end{bmatrix}, \qquad WB = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \qquad T^{-1}x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and let $\nu = \operatorname{ind}(E, A)$. We call (4.2a) the *slow subsystem* of dimension n_f and (4.2b) the *fast subsystem* of dimension n_∞. Then, we know that the state response of (4.2) is

$$x_1(t) = e^{Jt}x_1(0) + \int_o^t e^{J(t-s)}B_1u(s)ds \qquad (t > 0)$$

$$x_2(t) = -\sum_{i=0}^{\nu-1} N^iB_2u^{(i)}.$$

Thus, an admissible control function (for a classical solution) has to satisfy $u \in \mathcal{C}_p^{\nu-1}(\mathbb{I}, \mathbb{R}^m)$ (i.e. $(\nu-1)$- times piecewise continuously differentiable). For any $t > 0$ the state response

$x(t) = T \begin{bmatrix} x_1^\top & x_2^\top \end{bmatrix}^\top$ is uniquely determined by the initial condition $x_1(0)$, the control input $u(s)$, $0 \leq s \leq t$, and the time point t. In particular, the initial condition $x_2(0)$ has to be consistent (i.e. $x_2(0)$ is uniquely determined) and only $x_1(0)$ can be choosen arbitrarily. In the following, we denote by $\tilde{\mathcal{R}}_0$ the reachable set of (4.2) from the zero initial condition $x_1(0) = 0$ (and consistent $x_2(0)$).

Lemma 4.6. *For any polynomial $p(t) \not\equiv 0$ consider the matrix*

$$W(p,t) = \int_0^t p(s) e^{A_1 s} B_1 B_1^\top e^{A_1^\top s} p(s) ds,$$

with $A_1 \in \mathbb{R}^{n,n}, B_1 \in \mathbb{R}^{n,m}$. Then, it holds that

$$Im(W(p,t)) = Im \begin{bmatrix} B_1 & A_1 B_1 & \dots & A_1^{n-1} B_1 \end{bmatrix}$$

for any $t > 0$.

Proof. The statement of the lemma is equivalent to the statement that

$$\ker(W(p,t)) = \bigcap_{i=0}^{n-1} \ker \left(B_1^\top \left(A_1^\top \right)^i \right).$$

Let $x \in \ker(W(p,t))$, then

$$x^\top W(p,t) x = \int_0^t x^\top p(s) e^{A_1 s} B_1 B_1^\top e^{A_1^\top s} p(s) x \, ds = \int_0^t \underbrace{\| B_1^\top e^{A_1^\top s} p(s) x \|_2^2}_{\geq 0} ds = 0$$

and hence $B_1^\top e^{A_1^\top s} p(s) x = 0$ for $0 \leq s \leq t$. The polynomial $p(s)$ has a finite number of roots in $[0,t]$, thus we have

$$B_1^\top e^{A_1^\top s} x = 0, \qquad 0 \leq s \leq t.$$

Since s is arbitrary, we have $x \in \bigcap_{i=0}^{n-1} \ker \left(B_1^\top \left(A_1^\top \right)^i \right)$ (by Cayley-Hamilton) and thus $\ker(W(p,t)) \subseteq \ker \left(B_1^\top \left(A_1^\top \right)^i \right)$. For $x \in \ker \left(B_1^\top \left(A_1^\top \right)^i \right)$ the reverse of this process yields $x \in \ker(W(p,t))$, which implies

$$\ker \left(B_1^\top \left(A_1^\top \right)^i \right) \subseteq \ker(W(p,t)).$$

□

Lemma 4.7. *Let $x_i \in \mathbb{R}^n$, $i = 0, 1, \dots, \nu - 1$ and $t_1 > 0$. Then there exists a polynomial $p(t) \in \mathbb{R}^n$ of order $\nu - 1$ such that $p^{(i)}(t_1) = x_1$ for $i = 0, 1, \dots, \nu - 1$.*

Proof. Straightforward by setting

$$p(t) = x_0 + x_1(t-t_1) + \ldots + \frac{1}{(\nu-1)!}x_{\nu-1}(t-t_1)^{\nu-1}.$$

□

Theorem 4.8. *Let $\tilde{\mathcal{R}}_0$ be the reachable set of (4.2) from the zero initial condition $x_1(0) = 0$. Then*

$$\tilde{\mathcal{R}}_0 = Im\begin{bmatrix} B_1 & JB_1 & \ldots & J^{n_f-1}B_1 \end{bmatrix} \oplus Im\begin{bmatrix} B_2 & NB_2 & \ldots & N^{n_\infty-1}B_2 \end{bmatrix}.$$

Remark 4.9. In Theorem 4.8, $\oplus$ is meant as the cartesian product, i.e. $\oplus = \times$, as this is the common notation in the literature.

Proof. For $x_1(0) = 0$ and $t > 0$ the state response of (4.2) is given by

$$x_1(t) = \int_0^t e^{J(t-s)}B_1u(s)ds, \qquad x_2(t) = -\sum_{i=0}^{\nu-1} N^iB_2u^{(i)}(t).$$

Obviously, $x_2(t) \in \operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \ldots & N^{n_\infty-1}B_2 \end{bmatrix}$ (since $\nu \leq n_\infty$). Furthermore, there exists (exercise) analytic functions $\beta_i(t) \in \mathbb{R}$, $i = 0, \ldots, n_f - 1$ such that

$$e^{Jt} = \beta_0(t)I + \beta_1(t)J + \ldots + \beta_{n_f-1}(t)J^{n_f-1}$$

(see Exercise). Thus,

$$x_1(t) = \int_0^t e^{J(t-s)}B_1u(s)ds = \sum_{i=0}^{n_f-1} J^iB_1 \underbrace{\int_0^t \beta_i(t-s)u(s)ds}_{:=w(t)} \in \operatorname{Im}\begin{bmatrix} B_1 & JB_1 & \ldots & J^{n_f-1}B_1 \end{bmatrix}$$

for all $t > 0$. Hence

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \in \operatorname{Im}\begin{bmatrix} B_1 & JB_1 & \ldots & J^{n_f-1}B_1 \end{bmatrix} \oplus \operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \ldots & N^{n_\infty-1}B_2 \end{bmatrix}.$$

On the other hand, let

$$\hat{x} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} \in \operatorname{Im}\begin{bmatrix} B_1 & JB_1 & \ldots & J^{n_f-1}B_1 \end{bmatrix} \oplus \operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \ldots & N^{n_\infty-1}B_2 \end{bmatrix},$$

with $\hat{x}_1 \in \operatorname{Im}\begin{bmatrix} B_1 & JB_1 & \ldots & J^{n_f-1}B_1 \end{bmatrix}$ and $\hat{x}_2 \in \operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \ldots & N^{n_\infty-1}B_2 \end{bmatrix}$. Thus, there exists $w_i \in \mathbb{R}^{n_\infty}$, $i = 0, \ldots, \nu - 1$ such that

$$\hat{x}_2 = -\sum_{i=0}^{\nu-1} N^iB_2w_i.$$

From Lemma 4.7 for any fixed $t > 0$ there exists a polynomial $p(s)$ of order $\nu - 1$ such that $p^{(i)}(t) = w_i$. Thus, using the input function $u(t) = u_1(t) + p(t)$ we get the system response

$$x_1(t) = \int_0^t e^{J(t-s)} B_1 u_1(s) ds + \int_0^t e^{J(t-s)} B_1 p(s) ds$$

and

$$\tilde{x}_1 := \hat{x}_1 - \int_0^t e^{J(t-s)} B_1 p(s) ds \in \operatorname{Im} \begin{bmatrix} B_1 & JB_1 & \dots J^{n_f-1} B_1 \end{bmatrix}$$

for some fixed $t > 0$. For any fixed $t > 0$ let $q(s) = s^\nu (s-t)^\nu \neq 0$. From Lemma 4.6 we deduce the existence of $z \in \mathbb{R}^{n_f}$ such that $W(q,t)z = \tilde{x}_1$. Setting $u_1(s) = q(s)^2 B_1^\top e^{J^\top (t-s)} z$ for $0 \leq s \leq t$ we get the system response

$$\begin{aligned} x_1(t) &= \int_0^t e^{J(t-s)} B_1 q(s)^2 B_1^\top e^{J^\top (t-s)} z ds + \int_0^t e^{J(t-s)} B_1 p(s) ds \\ &= \int_0^t q(s) e^{J(t-s)} B_1 B_1^\top e^{J^\top (t-s)} q(s) ds z + \int_0^t e^{J(t-s)} B_1 p(s) ds \\ &= W(q,t)z + \hat{x}_1 - \tilde{x}_1 = \hat{x}_1 \end{aligned}$$

and

$$\begin{aligned} x_2(t) &= -\sum_{i=0}^{\nu-1} N^i B_2 u^{(i)}(t) = -\sum_{i=0}^{\nu-1} N^i B_2 (\underbrace{u_1(t)}_{=0} + \underbrace{p^{(i)}(t)}_{=w_i}) \\ &= -\sum_{i=0}^{\nu-1} N^i B_2 w_i = \hat{x}_2. \end{aligned}$$

Hence, $\hat{x} \in \tilde{\mathcal{R}}_0$ and the result follows. □

Example 4.10.

1. *Consider the system*

$$\begin{aligned} \dot{x}_1 &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \qquad x_1(0) = x_1^0, \\ 0 &= x_2 + \begin{bmatrix} -1 & 0 \end{bmatrix} u, \end{aligned}$$

with $n = 4, n_f = 2, n_\infty = 2$, which is already in Weierstraß canonical form (WCF). *The reachable set from $x_1(0) = 0$ is given by*

$$\tilde{\mathcal{R}}_0 = Im \begin{bmatrix} B_1 & JB_1 \end{bmatrix} \oplus \begin{bmatrix} B_2 & NB_2 \end{bmatrix} = \mathbb{R}^2 \oplus (\mathbb{R} \oplus \{0\}) = \mathbb{R}^3 \oplus \{0\}.$$

2. *Consider the system*

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \dot{x}_2 = x_2 + \begin{bmatrix} -1 \\ -1 \end{bmatrix} u.$$

The reachable set is given by $\tilde{\mathcal{R}}_0 = \mathbb{R}^2$ *and the state response of this system is given by*

$$x_2(t) = -\sum_{i=0}^{\nu-1} N^i B_2 u^{(i)}(t) = \begin{bmatrix} u + \dot{u} \\ u. \end{bmatrix}$$

Thus, for any $w = \begin{bmatrix} w_1 & w_2 \end{bmatrix}^\top \in \mathbb{R}^2$ *and* $t_1 > 0$ *we can choose* $u(t)$ *such that* $u(t_1) = w_2$, $\dot{u}(t_1) = w_1 - w_2$ *and we get*

$$x_2(t_1) = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

Note that such a control strategy may require high input energy, since for large $\|w_1 - w_2\|$ $\dot{u}$ *will also be large.*

For the next results, we need the following reminder from control theory.

Lemma 4.11 (Hautus-Popov-Lemma)**.** *The following are equivalent:*

1. *The system* $\dot{x} = Ax + Bu$ *is C-controllable.*
2. *rank*$(K) = rank \begin{bmatrix} B & AB & \dots & A^{n-1}B \end{bmatrix} = n.$
3. *If* z *is eigenvector of* $A^\top$*, then* $z^\top B \neq 0.$
4. *rank* $\begin{bmatrix} \lambda I - A & B \end{bmatrix} = n$ *for all* $\lambda \in \mathbb{C}.$

Theorem 4.12.

1. *The slow subsystem* (4.2a) *is C-controllable if and only if*

$$rank \begin{bmatrix} \lambda E - A & B \end{bmatrix} = n \qquad \text{for all } \lambda \in \mathbb{C}, \lambda \text{ finite.}$$

2. *The following statements are equivalent:*

   a) *The fast subsystem* (4.2b) *is C-controllable.*

   b) *rank* $\begin{bmatrix} B_2 & NB_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix} = n_\infty.$

   c) *rank* $\begin{bmatrix} N & B_2 \end{bmatrix} = n_\infty.$

   d) *rank* $\begin{bmatrix} E & B \end{bmatrix} = n.$

   e) *For any nonsingular matrices* Q_1 *and* P_1 *satisfying*

$$E = Q_1 \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} P_1, \qquad \text{let} \quad QB = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix}.$$

Then $\tilde{B}_2$ is of full row rank, rank$(\tilde{B}_2) = n - $ rank(E).

3. *The following statements are equivalent:*

   a) *The system* (4.2) *is C-controllable.*

   b) *The slow and fast subsystems* (4.2a) *and* (4.2b) *are both C-controllable.*

   c) *rank* $\begin{bmatrix} B_1 & JB_1 & \cdots & J^{n_f-1}B_1 \end{bmatrix} = n_f$ *and rank* $\begin{bmatrix} B_2 & NB_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix} = n_\infty$.

   d) *rank* $\begin{bmatrix} \lambda E - A & B \end{bmatrix} = n$ *for all finite* $\lambda \in \mathbb{C}$ *and rank* $\begin{bmatrix} E & B \end{bmatrix} = n$.

   e) *rank* $\begin{bmatrix} \alpha E - \beta A & B \end{bmatrix} = n$ *for all* $(\alpha, \beta) \in \mathbb{C}^2 \backslash \{(0,0)\}$.

Proof.

1. The slow subsystem is an ODE, thus the controllability conditions for standard LTI systems apply and (4.2a) is C-controllable if and only if rank $\begin{bmatrix} \lambda I - J & B_1 \end{bmatrix} = n_f$ for all finite $\lambda \in \mathbb{C}$. Furthermore, we have

$$\operatorname{rank} \begin{bmatrix} \lambda E - A & B \end{bmatrix} \overset{\text{(WCF)}}{=} \operatorname{rank} \begin{bmatrix} \lambda WET - WAT & WB \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \lambda I - J & 0 & B_1 \\ 0 & \lambda N - I & B_2 \end{bmatrix}.$$

   The matrix $\lambda N - I$ is nonsingular for any finite $\lambda \in \mathbb{C}$ and hence

$$\operatorname{rank} \begin{bmatrix} \lambda E - A & B \end{bmatrix} = n_\infty + \operatorname{rank} \begin{bmatrix} \lambda I - J & B_1 \end{bmatrix} = n.$$

2. **2a** $\iff$ **2b** By definition the fast subsystem (4.2b) is C-controllable if the reachable set is

$$\operatorname{Im} \begin{bmatrix} B_2 & NB_2 & \dots N^{\nu-1}B_2 \end{bmatrix} = \mathbb{R}^{n_\infty} \iff \operatorname{rank} \begin{bmatrix} B_2 & NB_2 & \dots N^{\nu-1}B_2 \end{bmatrix} = n_\infty.$$

   2b $\iff$ **2c** The system (N, B_2) is C-controllable (as a standard LTI system) if and only if rank $\begin{bmatrix} \lambda I - N, B_2 \end{bmatrix} = n_\infty$ for all $\lambda \in \mathbb{C}$. Thus, this condition holds for all $\lambda \in \sigma(N) = \{0\}$ since N nilpotent and thus

$$\operatorname{rank} \begin{bmatrix} \lambda I - N & B_2 \end{bmatrix} = n_\infty \text{ for all } \lambda \in \mathbb{C} \iff \operatorname{rank} \begin{bmatrix} -N & B_2 \end{bmatrix} = \operatorname{rank} \begin{bmatrix} N & B_2 \end{bmatrix} = n_\infty.$$

   2c $\iff$ **2d** We have

$$\operatorname{rank} \begin{bmatrix} E & B \end{bmatrix} = \operatorname{rank} \begin{bmatrix} WET & WB \end{bmatrix} = \operatorname{rank} \begin{bmatrix} I_{n_f} & 0 & B_1 \\ 0 & N & B_2 \end{bmatrix} = n_f + \operatorname{rank} \begin{bmatrix} N & B_2 \end{bmatrix}.$$

   Thus, rank $\begin{bmatrix} N & B_2 \end{bmatrix} = n_\infty \iff$ rank $\begin{bmatrix} E & B \end{bmatrix} = n$.

2d $\iff$ **2e** Similar as the previous statement.

3. **3a** $\iff$ **3c** Let the system be C-controllable and let $x_1(0) = 0$. Then for any $t_1 > 0$ and $w \in \mathbb{R}^n$, there exists an admissible control input $u \in \mathcal{C}_p^{\nu-1}$ such that $x(t_1) = w$. Thus

$$\begin{aligned}&\tilde{\mathcal{R}}_0 = \operatorname{Im}\begin{bmatrix}B_1 & JB_1 & \cdots & J^{n_f-1}B_1\end{bmatrix} \oplus \operatorname{Im}\begin{bmatrix}B_2 & NB_2 & \cdots N^{\nu-1}B_2\end{bmatrix} = \mathbb{R}^n\\ \iff &\operatorname{rank}\begin{bmatrix}B_1 & JB_1 & \cdots & J^{n_f-1}B_1\end{bmatrix} = n_f \text{ and } \operatorname{rank}\begin{bmatrix}B_2 & NB_2 & \cdots N^{\nu-1}B_2\end{bmatrix} = n_\infty.\end{aligned}$$

On the other hand, let the rank conditions hold. Then, we know that

$$\mathcal{R}_{x_1(0)} = \tilde{\mathcal{R}}_0 + \left\{ \begin{bmatrix}x_1\\x_2\end{bmatrix} \mid x_1 = e^{Jt}x_1(0) \in \mathbb{R}^{n_f}, x_2 = 0 \in \mathbb{R}^{n_\infty} \right\} = \mathbb{R}^n$$

and thus the system (4.2) is C-controllable.

3b $\iff$ **3c** Follows from 1 and 2.

3b $\iff$ **3d** Clear from 1 and 2.

3d $\iff$ **3e** $\operatorname{rank}\begin{bmatrix}\lambda E - A & B\end{bmatrix} = \operatorname{rank}\begin{bmatrix}\frac{\alpha}{\beta}E - A & B\end{bmatrix} = \operatorname{rank}\begin{bmatrix}\alpha E - \beta A & B\end{bmatrix}$.

□

Example 4.13. *Consider the system*

$$\begin{aligned}\dot{x}_1 &= \begin{bmatrix}1 & 1\\0 & 1\end{bmatrix} x_1 + \begin{bmatrix}0\\1\end{bmatrix} u\\ 0 &= x_2 + \begin{bmatrix}-1\\0\end{bmatrix} u.\end{aligned}$$

We have $\operatorname{rank}\begin{bmatrix}B_1 & JB_1\end{bmatrix} = \operatorname{rank}\begin{bmatrix}0 & 1\\1 & 1\end{bmatrix} = 2$ *and* $\operatorname{rank}\begin{bmatrix}B_2 & BN_2\end{bmatrix} = \operatorname{rank}\begin{bmatrix}-1 & 0\\0 & 0\end{bmatrix} = 1 < 2$. *Thus, the system is not C-controllable, while the slow subsystem is C-controllable.*

Remark 4.14. For systems with large state dimension n, the criteria given in the previous theorem are not suitable for numerical computations, since the system decomposition into (WCF) or the eigenvalues are needed. A better way for numerics is via staircase forms (see chapter 5).

A regular linear time-invariant descriptor system is R-controllable (i.e. controllable within the reachable set) if for any given $t_f > 0$ and $x_1(0) \in \mathcal{R}$, $w \in \mathcal{R}$, there exists an admissible control input $u \in \mathcal{C}_p^{\nu-1}$ such that $x(t_f) = w$.

Theorem 4.15. *The following statements are equivalent:*

1. *The system* (4.2) *is R-controllable.*

2. *The slow subsystem is C-controllable.*

3. *rank* $\begin{bmatrix}\lambda E - A & B\end{bmatrix} = n$ *for all finite* $\lambda \in \mathbb{C}$.

4. *rank* $\begin{bmatrix}B_1 & JB_1 & \cdots & J^{n_f-1}B_1\end{bmatrix} = n_f$.

Proof.

1 $\iff$ **1** By definition the system is R-controllable if

$$\begin{aligned}\tilde{\mathcal{R}}_0 &= \operatorname{Im}\begin{bmatrix}B_1 & JB_1 & \cdots & J^{n_f-1}B_1\end{bmatrix} \oplus \operatorname{Im}\begin{bmatrix}B_2 & NB_2 & \cdots N^{\nu-1}B_2\end{bmatrix} \\ &= \mathbb{R}^{n_f} \oplus \operatorname{Im}\begin{bmatrix}B_2 & NB_2 & \cdots & N^{\nu-1}B_2\end{bmatrix}.\end{aligned}$$

Thus, $\operatorname{Im}\begin{bmatrix}B_1 & JB_1 & \cdots J^{n_f-1}B_1\end{bmatrix} = \mathbb{R}^{n_f} \iff$ the slow subsystem (4.2a) is C-controllable.

2 $\iff$ **3** follows directly from Theorem 4.12.

□

Thus, C-controllability implies R-controllability. The converse direction is not true in general.

Example 4.16.

1. *Consider the system from Example 4.13 given by*

$$\begin{aligned}\dot{x}_1 &= \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix} x_1 + \begin{bmatrix}0\\1\end{bmatrix} u \\ 0 &= x_2 + \begin{bmatrix}-1\\0\end{bmatrix} u.\end{aligned}$$

The slow subsystem is C-controllable as we have already seen and hence the system is R-controllable.

2. *Let* N *be nilpotent and consider the system* $N\dot{x} = x + Bu$. *The system consists only of the fast subsystem and this is always R-controllable.*

Corollary 4.17. *Consider system* (4.1) *with regular matrix pencil* $\lambda E - A$. *Then, the system is C-controllable if and only if the system is R-controllable and rank* $\begin{bmatrix}E & B\end{bmatrix} = n$.

Proof. Clear from previous results. □

Exkurs: Generalized Functions and Distributional Solutions

We want to allow for jumps/discontinuities in the solution $x(t)$ at a number of distinct points in $\mathbb{I}$ in a distributional setting. In the following, let $\mathcal{D}^n = \mathcal{C}_0^\infty(\mathbb{R}, \mathbb{R}^n)$ denote the set of infinitely differentiable functions with values in $\mathbb{R}^n$ and compact support in $\mathbb{R}$. The elements of $\mathcal{D}^n$ are called *test functions*.

Definition 4.18. A linear functional $f : \mathcal{D}^n \to \mathbb{R}^n$ with

$$f(\alpha_1\phi_1 + \alpha_2\phi_2) = \alpha_1 f(\phi_1) + \alpha_2 f(\phi_2) \qquad \text{for all } \phi_1, \phi_2 \in \mathcal{D}^n, \alpha_1, \alpha_2 \in \mathbb{R}$$

is called a *generalized function* or a *distribution* if it is continuous, that is $f(\phi) \to 0$ in $\mathbb{R}^n$ for all sequences $(\phi_i)_{i\in\mathbb{N}}$ with $\phi_i \to 0$ in $\mathcal{D}^n$. [A sequence $(\phi_i(t))_{i\in\mathbb{N}}$ converges to zero if all functions ϕ_i vanish outside a bounded interval and $\left(\phi_i^{(q)}(t)\right)_{i\in\mathbb{N}}$ converges uniformly to zero for all $q \in \mathbb{N}_0$.] We denote the *space of all distributions* acting on $\mathcal{D}^n$ by $\mathcal{C}^n$.

In order to use distributions in the framework of descriptor systems we need the notion of derivatives and primitives of distributions. The q-th order derivative $f^{(q)}$, $q \in \mathbb{N}_0$ of a distribution $f \in \mathcal{C}^n$ is defined by

$$f^{(q)}(\phi) = (-1)^q f(\phi^{(q)}) \qquad \text{for all } \phi \in \mathcal{D}^n.$$

The functional $f^{(q)}$ is linear and continuous, thus every distribution has derivatives of arbitrary order in $\mathcal{C}^n$. For a distribution $f \in \mathcal{C}^n$ any distribution $X \in \mathcal{C}^n$ that satisfies

$$\dot{X}(\phi) = f(\phi) \qquad \text{for all } \phi \in \mathcal{D}^n$$

is called a primitive of f, i.e. X is a solution of $\dot{X} = f$. For $A \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{m,n})$ and $x \in \mathcal{C}^n$ multiplication by a matrix-valued function is defined by

$$Ax(\phi) = x(A^\top \phi) \qquad \text{for all } \phi \in \mathcal{D}^n.$$

Example 4.19. *The* Dirac delta distribution $\delta_\alpha \in \mathcal{C}^n$ *is defined by* $\delta_\alpha(\phi) = \phi(\alpha)$ *for all* $\phi \in \mathcal{D}^n$, $\alpha \in \mathbb{R}$. *It can be loosely thought of as*

$$\delta_\alpha(x) = \begin{cases} +\infty & x = \alpha, \\ 0 & \textit{otherwise.} \end{cases}$$

Remark 4.20. Since for a given $\phi \in \mathcal{D}^n$ and $\hat{t} > 0$ sufficiently large it holds that

$$\phi(0) = -(\phi(\hat{t}) - \phi(0)) = -\phi(t)\big|_0^{\hat{t}} = -\int_0^{\hat{t}} \dot{\phi}(t)dt = -\int_0^{\infty} \dot{\phi}(t)dt = -\int_{\mathbb{R}} H(t)\dot{\phi}(t)dt =: H(\dot{\phi}),$$

where $H(t) = \begin{cases} 0 & \text{for } t < 0, \\ 1 & \text{for } t \geq 0 \end{cases}$ is the *Heaviside function*. We find the relation $\delta_0 = \dot{H}$. We can also define the shifted version of H by $H_\alpha(t) := H(t - \alpha)$ and $\delta_\alpha = \dot{H}_\alpha$.

Two distributions $f_1, f_2 \in \mathcal{C}^n$ are equal if $f_1(\phi) = f_2(\phi)$ for all $\phi \in \mathcal{D}^n$. In the following, a function $x : \mathbb{I} \to \mathbb{R}^n$, $\mathbb{I} \subseteq \mathbb{R}$ is treated as a function defined on $\mathbb{R}$ by setting $x(t) = 0$ for $t \notin \mathbb{I}$. Futhermore, nonsmooth behavior of the solution is restricted to happen at a countable number of points $\tau_j \in \mathbb{T} \subseteq \mathbb{R}$.

Definition 4.21. Suppose that the set $\mathbb{T} = \{\tau_j \in \mathbb{R} \big| \ \tau_j < \tau_{j+1}, j \in \mathbb{Z}\}$ has no accumulation points. A generalized function/distribution $x \in \mathcal{C}^n$ is called *impulsive smooth* if it can be written in the form

$$x = \hat{x} + x_{\text{imp}}, \qquad \hat{x} = \sum_{j \in \mathbb{Z}} \hat{x}_j, \tag{4.3}$$

where $\hat{x}_j \in \mathcal{C}^\infty([\tau_j, \tau_{j+1}], \mathbb{R}^n)$ for all $j \in \mathbb{Z}$ and the impulsive part x_{imp} has the form

$$x_{\text{imp},j} = \sum_{i=0}^{q_j} c_{ij} \delta_{\tau_j}^{(i)}, \qquad c_{ij} \in \mathbb{R}^n, q_j \in \mathbb{N}_0. \tag{4.4}$$

The set of all impulsive smooth distributions is denoted by $\mathcal{C}^n_{\text{imp}}(\mathbb{T})$.

Lemma 4.22.

1. *A distribution $x \in \mathcal{C}^n_{imp}(\mathbb{T})$ uniquely determines the decomposition* (4.3).
2. *For $x \in \mathcal{C}^n_{imp}(\mathbb{T})$ we can asign a function value $x(t)$ for every $t \in \mathbb{R}\backslash\mathbb{T}$ by $x(t) = \hat{x}_j(t)$ for $t \in (\tau_j, \tau_{j+1})$ and limits $x(\tau_j^-) = \lim_{t \to \tau_j^-} \hat{x}_{j-1}(t)$ and $x(\tau_j^+) = \lim_{t \to \tau_j^+} \hat{x}_j(t)$ for every $\tau_j \in \mathbb{T}$.*
3. *All derivatives and primitives of $x \in \mathcal{C}^n_{imp}(\mathbb{T})$ are again in $\mathcal{C}^n_{imp}(\mathbb{T})$.*
4. *The set $\mathcal{C}^n_{imp}(\mathbb{T})$ is a vector space and closed under multiplication with functions $A \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{m,n})$.*

Proof. See [6, Lemma 3.75]. □

We can also introduce a measure for the smoothness of impulsive smooth distributions as follows.

Definition 4.23. The *impulse order* of $x \in \mathcal{C}^n_{\text{imp}}(\mathbb{T})$ at $\tau_j \in \mathbb{T}$ is defined as $\operatorname{iord}(x)\big|_{\tau_j} := -q-2$ if x can be associated with a continuous function in $[\tau_{j-1}, \tau_{j+1}]$ and q, with $0 \leq q \leq \infty$ is the largest integer such that

$$x\big|_{[\tau_{j-1}, \tau_{j+1}]} \in \mathcal{C}^q([\tau_{j-1}, \tau_{j+1}], \mathbb{R}^n).$$

It is defined as $\operatorname{iord}(x)\big|_{\tau_j} := -1$ if x can be associated with a function that is continuous in $[\tau_{j-1}, \tau_{j+1}]$ except at $t = \tau_j$ and it is defined as

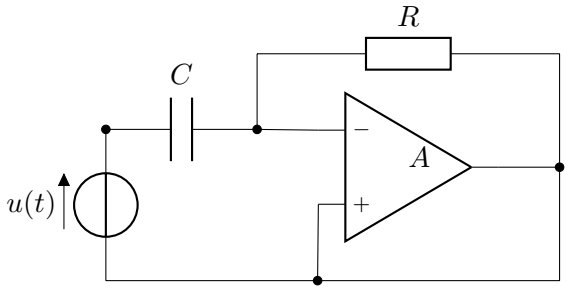
$$\operatorname{iord}(x)\big|_{\tau_j} := \max\{i \in \mathbb{N}_0 \big| \ 0 \leq i \leq q_j, c_{ij} \neq 0\}$$

otherwise. The impulse order of x is defined as $\operatorname{iord} x := \max_{\tau_j \in \mathbb{T}} \operatorname{iord}(x)\big|_{\tau_j}$.

Lemma 4.24. *Let $x \in \mathcal{C}^n_{imp}(\mathbb{T})$ and $A \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{m,n})$. Then iord $Ax \leq$ iord x with equality for $m = n$ and $A(\tau_j)$ invertible for each $\tau_j \in \mathbb{T}$.*

Proof. See [6]. □

Example 4.25. *Consider the model of a differentiator circuit (see Figure 4.2) described by the following DAE*



C: linear capacitor
R: linear resistor
A: linear amplifier
x_i: node potential at node i

Figure 4.2: Model of a differentiator circuit

$$\begin{aligned} x_1 - x_4 &= u(t), \\ C(\dot{x}_1 - \dot{x}_2) + \frac{1}{R}(x_3 - x_2) &= 0, \\ x_3 &= A(x_4 - x_2), \\ x_4 &= 0, \end{aligned}$$

with input voltage $u(t) = \begin{cases} 0 & \text{for } t < 0, \\ 1 & \text{for } t \geq 0. \end{cases}$

With $x_4 = 0$, $x_1 = u(t)$, $\dot{x}_1 = \dot{u}$ and $x_3 = -Ax_2$ we get

$$\dot{x}_2 = -\frac{1}{CR}(A+1)x_2 + \dot{u}.$$

Due to the jump in the input voltage, u is not differentiable. For $u = H$ and $A \to \infty$, the model equations take the form

$$\begin{aligned} x_1 - x_4 &= H, \\ C(\dot{x}_1 - \dot{x}_2) + \frac{1}{R}(x_3 - x_2) &= 0, \\ x_2 &= 0, \\ x_4 &= 0, \end{aligned}$$

with solution $x_1 = H$, $x_2 = 0$, $x_3 = -RC\dot{H} = -RC\delta_0$ and $x_4 = 0$. The system is a DAE of index $\nu = 2$ (or $\mu = 1$) and iord $f = -1$. For a consistent initial value, e.g. $x(-1) = 0$ we have a unique solution x with iord $x = 0$.

For the special case of regular time-invariant DAEs of the form $E\dot{x} = Ax + f$ with $f \in \mathcal{C}^n_{\mathrm{imp}}(\mathbb{T})$ of impulse order $\operatorname{iord} f = q \in \mathbb{Z} \cup \{-\infty\}$ we can proceed as follows. First, we can transform the matrix pair (E, A) into (WCF)

$$(E, A) \sim (WET, WAT) = \left(\begin{bmatrix} I_{n_f} & 0 \\ 0 & N \end{bmatrix}, \begin{bmatrix} J & 0 \\ 0 & I_{n_\infty} \end{bmatrix} \right).$$

Thus, we get

$$\dot{x}_1 = Jx_1 + f_1, \tag{4.5a}$$
$$N\dot{x}_2 = x_2 + f_2, \tag{4.5b}$$

where $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = T^{-1}x$. For the distributional ODE (4.5a) we can consider the fundamental solution matrix $X(t)$ that satisfies

$$\dot{X}(t) = JX(t), \qquad X(t_0) = I,$$

i.e. $X(t) = e^{J(t-t_0)} \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{n_f, n_f})$. Thus, a distribution $\tilde{x} \in \mathcal{C}^n_{\mathrm{imp}}(\mathbb{T})$ solves (4.5a) if and only if $z = X^{-1}\tilde{x} \in \mathcal{C}^n_{\mathrm{imp}}(\mathbb{T})$ solves

$$\dot{z} = g_1 = X^{-1} f_1,$$

i.e. z is a primitive of g_1. Since $f \in \mathcal{C}^n_{\mathrm{imp}}(\mathbb{T})$ we have $f_1 \in \mathcal{C}^{n_f}_{\mathrm{imp}}(\mathbb{T})$ and hence $g_1 \in \mathcal{C}^{n_f}_{\mathrm{imp}}(\mathbb{T})$ with $\operatorname{iord} g_1 = \operatorname{iord} f_1$ by Lemma 4.24. Using the decomposition (4.3) given by $g_1 = \hat{g}_1 + g_{1,\mathrm{imp}}$ with

$$\hat{g}_1 = \sum_{j \in \mathbb{Z}} \hat{g}_{1,j} \qquad \text{and} \qquad g_{1,\mathrm{imp}} = \sum_{j \in \mathbb{Z}} \sum_{i=0}^{q_i} c_{ij} \delta_{\tau_j}^{(i)}$$

with the convention that $g_{1,\mathrm{imp},j} = \sum_{i=0}^{q_j} c_{ij} \delta_{\tau_j}^{(i)} = 0$ if $q_j < 0$, a primitive of g_1 has the form

$$\begin{aligned} z &= c + \int_{t_0}^{t} \sum_{j \in \mathbb{Z}} \hat{g}_{1,j}(s) ds + \int_{t_0}^{t} \sum_{j \in \mathbb{Z}} \sum_{i=0}^{q_i} c_{ij} \delta_{\tau_j}^{(i)} \\ &= c + \int_{t_0}^{t} \sum_{j \in \mathbb{Z}} \hat{g}_{1,j}(s) ds + \sum_{j \in \mathbb{Z}} \left(\int_{t_0}^{t} c_{0j} \delta_{\tau_j} + \sum_{i=1}^{q_i} c_{ij} \delta_{\tau_j}^{(i)} \right) \\ &= c + \int_{t_0}^{t} \sum_{j \in \mathbb{Z}} \hat{g}_{1,j}(s) ds + \sum_{j \in \mathbb{Z}} c_{0j} H_{\tau_j} + \sum_{j \in \mathbb{Z}} \sum_{i=0}^{q_i - 1} c_{ij} \delta_{\tau_j}^{(i)} \end{aligned}$$

with some $c \in \mathbb{R}^{n_f}$. Hence, every primitive z of g_1 has impulse order $q - 1$ and so every solution of (4.5a). For the initial value it holds that

$$z(t_0) = c + \sum_{j \in \mathbb{Z}} c_{0j} H_{\tau_j}(t_0) \qquad \text{for } t_0 \in \mathbb{R} \backslash \mathbb{T}$$

and

$$z(\tau_i^-) = c + \sum_{j\in\mathbb{Z}} c_{0j} \lim_{t\to\tau_i^-} H_{\tau_j}(t) = c + \sum_{j\in\mathbb{Z},\tau_j>\tau_i} c_{0j},$$
$$z(\tau_i^+) = c + \sum_{j\in\mathbb{Z}} c_{0j} \lim_{t\to\tau_i^+} H_{\tau_j}(t) = c + \sum_{j\in\mathbb{Z},\tau_j\geq\tau_i} c_{0j}.$$

Thus, there exists a unique solution of (4.5a) in $\mathcal{C}^n_{\text{imp}}(\mathbb{T})$ satisfying one of the initial conditions

$$x_1(t_0) = x_{1,0}, \qquad x_1(\tau_i^-) = x_{1,0}, \qquad x_1(\tau_i^+) = x_{1,0}$$

with $\tau_i \in \mathbb{T}$. For the algebraic part (4.5b) we use

$$f_2 = \hat{f}_2 + f_{2,\text{imp}} \in \mathcal{C}^{n\infty}_{\text{imp}}(\mathbb{T}).$$

The unique solution of (4.5b) is given by

$$x_2 = -\sum_{i=0}^{\nu-1} N^i \left(\hat{f}_2^{(i)} + f_{2,\text{imp}}^{(i)}\right) \in \mathcal{C}^{n\infty}_{\text{imp}}(\mathbb{T})$$

with $\operatorname{iord}(x_2) \leq q + \nu - 1$. Similar as before, consistency of an initial value implies that $x_2(t_0) = x_{2,0}$ for $t_0 \in \mathbb{R}\backslash\mathbb{T}$.

Theorem 4.26. *Let (E, A) be regular with $\nu = ind(E, A)$. Let $x_0^j \in \mathcal{C}^\infty([\tau_{j-1}, \tau_j], \mathbb{R}^n)$ be given and let $f = \hat{f} + f_{imp}$, $\hat{f} = \sum_{i\in\mathbb{Z}} \hat{f}_i$, where $\hat{f}_j = E\dot{x}_0^j - Ax_0^j$. Then, the following statements hold:*

1. *The DAE*
$$E\dot{x} = Ax + f \qquad \text{with } \hat{x}_{j-1} = x_0^j \tag{4.6}$$
has a unique solution $x \in \mathcal{C}^n_{imp}$ with $\operatorname{iord} x \leq \operatorname{iord} f + \nu - 1$.
2. *Let $x = \hat{x} + x_{imp}$, $\hat{x} = \sum_{i\in\mathbb{Z}} \hat{x}_i$ be the unique solution of (4.6). Then $\tilde{x} = x - \hat{x}_{j-1}$ is the unique solution of*
$$E\dot{\tilde{x}} = A\tilde{x} + \tilde{f} + Ex_{j,0}\delta_{\tau_j}, \qquad \tilde{x}_{j-1} = 0,$$
where $x_{j,0} = \hat{x}_{j-1}(\tau_j)$ and $\tilde{f} = f - \hat{f}_j$.

Proof. See [6]. □

Remark 4.27. Setting $\hat{f}_j = E\dot{x}_0^j - Ax_0^j$ forces x_0^j to be a solution of

$$E\dot{x} = Ax + \hat{f}_j \qquad \text{for } t \in [\tau_j, \tau_{j+1}] \text{ for all } j \in \mathbb{Z}. \tag{4.7}$$

Thus, the inhomogeneity $\hat{f}_j$ is modified in such a way that a given initial condition $x(\tau_j^-) = x_{j,0}(\tau_j)$ is made consistent for (4.7). Since $\dot{x} = \dot{\hat{x}} + \dot{x}_{\text{imp}} + \sum (\hat{x}_i(\tau_i) - \hat{x}_{i-1}(\tau_i))\, \delta_{\tau_i}$ we have

$$\begin{aligned} E\left(\dot{\hat{x}} + \dot{x}_{\text{imp}} + \sum (\hat{x}_i(\tau_i) - \hat{x}_{i-1}(\tau_i))\, \delta_{\tau_i}\right) &= A\,(\hat{x} + x_{\text{imp}}) + \hat{f} + f_{\text{imp}} \\ &= A\,(\hat{x} + x_{\text{imp}}) + \sum_{j\in\mathbb{Z}, j\neq i} \hat{f}_i + f_{\text{imp}} + E\dot{x}_0^j - Ax_0^j. \end{aligned}$$

Setting $\tilde{x} = x - \hat{x}_{j-1}$ and $\tilde{f} = f - \hat{f}_j$ this can be expressed as

$$E\dot{\tilde{x}} = A\tilde{x} + \tilde{f} + Ex_{j,0}\delta_{\tau_j}, \qquad \tilde{x}_{j-1} = 0,$$

where $x_{j,0} = \hat{x}_{j-1}(\tau_j)$.

In the descriptor setting we consider

$$E\dot{x} = Ax + Bu \tag{4.8}$$

with input function $u \in \mathcal{C}_p^\infty(\mathbb{R}, \mathbb{R}^n)$ (infinitely often piecewise continuously differentiable). Thus, $u \in \mathcal{C}_{\text{imp}}^n(\mathbb{T})$ with $\operatorname{iord}(u) = q \leq -1$ and we can consider (4.8) in a distributional framework.

Remark 4.28. If we consider piecewise continuous distributions, then $u \in \mathcal{C}_p^{\nu-1}$ is sufficient (see [4]).

The corresponding system in (WCF) has the form

$$\dot{x}_1 = Jx_1 + B_1u. \tag{4.9a}$$

$$N\dot{x}_2 = x_2 + B_2u. \tag{4.9b}$$

Since $\operatorname{iord}(u) = q \leq -1$ we have that $\operatorname{iord}(x_1) = q - 1 \leq -1$ and thus (at least) $x_1 \in \mathcal{C}^0(\mathbb{R}, \mathbb{R}^n)$, i.e. there are no impulse terms in the system response x_1. The system response of the algebraic part (fast subsystem) takes the form

$$x_2 = -\sum_{i=0}^{\nu-1} N^i B_2 u^{(i)}$$

with $\operatorname{iord}(x_2) \leq q + \nu - 1 \leq \nu - 2$. From Theorem 4.26 we get that $\tilde{x}_2 = x_2 - \hat{x}_{2,j-1}$ is the unique solution of

$$N\dot{\tilde{x}}_2 = \tilde{x}_2 + B_2\tilde{u} + Nx_{2j,0}\delta_{\tau_j}, \qquad \tilde{x}_{2,j-1} = 0,$$

where $x_{2j,0} = \hat{x}_{2,j-1}(\tau_j)$ and $\tilde{u} = u - \hat{u}_j$. Thus,

$$\tilde{x}_2 = -\sum_{i=0}^{\nu-1} N^i B_2\tilde{u}^{(i)} - \sum_{i=0}^{\nu-1} N^{i+1}x_{2j,0}\delta_{\tau_j}^{(i)} = \underbrace{-\sum_{i=0}^{\nu-1} N^i B_2\tilde{u}^{(i)}}_{=P_2} \underbrace{-\sum_{i=1}^{\nu-1} N^i x_{2j,0}\delta_{\tau_j}^{(i-1)}}_{=P_1}.$$

There exist impulsive terms in the state response x_2 either due to the initial condition (P_1) or due to possible jump behaviors in $\tilde{u}$ and $\tilde{u}^{(i)}$ (P_2).

Theorem 4.29. *Consider* (4.8) *with regular pair* (E, A)*. Then there exists* $u \in \mathcal{C}_p^\infty$ *such that* $\Delta_{\tau_j} x := x(\tau_j^+) - x(\tau_j^-) = w$ *(the* jump *in* x *at* τ_j*) for some* $\tau_j \in \mathbb{T}$ *if and only if* $w \in 0 \oplus Im \begin{bmatrix} B_2 & NB_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix}$.

Proof.
"$\Rightarrow$" $\Delta_{\tau_j} x = \begin{bmatrix} 0 \\ \Delta_{\tau_j} x_2 \end{bmatrix}$ since $\Delta_{\tau_j} x_1 = 0$ for $u \in \mathcal{C}_p^\infty$, $\tau_j \in \mathbb{T}$. Thus,

$$\Delta_{\tau_j} x_2 = x_2(\tau_j^+) - x_2(\tau_j^-) = -\sum_{i=0}^{\nu-1} N^i B_2 \left(u^{(i)}(\tau_j^+) - u^{(i)}(\tau_j^-) \right).$$

Hence, $\Delta_{\tau_j} x_2 \in \operatorname{Im} \begin{bmatrix} B_2 & NB_2 & \cdots N^{\nu-1}B_2 \end{bmatrix}$.

"$\Leftarrow$" Choose $w_0, \ldots, w_{\nu-1}$ such that $w = -\sum_{i=0}^{\nu-1} N^i B_2 w_i$. Thus, we can choose

$$u(t) = \begin{cases} w_0 + (t-\tau_j)w_1 + \frac{1}{2}(t-\tau_j)^2 w_2 + \ldots + \frac{1}{(\nu-1)!}(t-\tau_j)^{\nu-1} w_{\nu-1} & \text{if } t \geq \tau_j, \\ 0 & \text{if } t < \tau_j. \end{cases}$$

and $\Delta_{\tau_j} x_2 = -\sum_{i=0}^{\nu-1} N^i B_2 \Delta_{\tau_j} u^{(i)} = -\sum_{i=0}^{\nu-1} N^i B_2 w_i = w$.

□

We define the mapping $\mathcal{I}_{\tau_j} : \mathbb{R}^{n_\infty} \to \mathcal{C}_{\text{imp}}^n$ by $\mathcal{I}_{\tau_j}(w) := \begin{bmatrix} 0 \\ \mathcal{I}_{2,\tau_j}(w) \end{bmatrix}$ with

$$\mathcal{I}_{2,\tau_j}(w) := -\sum_{i=0}^{\nu-1} \delta_{\tau_j}^{(i-1)} N^i w \qquad \text{for } \tau_j \in \mathbb{T}.$$

Note that $\mathcal{I}_{\tau=0}(x_2(0))$ represents the impulse behavior in $x(t)$ at the initial time point $t_0 = 0$ caused by the initial condition $x_2(0)$, and $\mathcal{I}_{\tau_j}(w)$ includes all possible impulse terms in $x(t)$ at τ_j.

Theorem 4.30. *Consider* (4.8) *with regular pair* (E, A)*. For any* $w \in \mathbb{R}^{n_\infty}$ *there exists* $u \in \mathcal{C}_p^\infty$ *such that the impulsive part of* x *at* τ_j *denoted by* $x_{imp,j}$ *is given by* $x_{imp,j} = \mathcal{I}_{\tau_j}(w)$ *if and only if*

$$\mathcal{I}_{2,\tau_j}(w) \in \begin{bmatrix} NB_2 & N^2B_2 & \cdots N^{\nu-1}B_2 \end{bmatrix},$$

i. e. $\left(\mathcal{I}_{2,\tau_j}(w)\right)(\phi) \in \begin{bmatrix} NB_2 & N^2B_2 & \cdots N^{\nu-1}B_2 \end{bmatrix}$ *for all* $\phi \in \mathcal{D}^n$.

Proof. Since $x_{1,\text{imp}} = 0$ we have $x_{\text{imp}} = \begin{bmatrix} 0 \\ x_{2,\text{imp}} \end{bmatrix}$ and $x_{\text{imp},j} = \begin{bmatrix} 0 \\ x_{2,\text{imp},j} \end{bmatrix}$. Similar as before using Theorem 4.26

$$\tilde{x}_2 = -\sum_{i=0}^{\nu-1} N^i B_2 \tilde{u}^{(i)} - \sum_{i=1}^{\nu-1} N^i x_{2j,0} \delta_{\tau_j}^{(i-1)}$$

and $\operatorname{iord}(\tilde{x}_2) \leq \nu - 2$ and hence

$$\tilde{x}_{2,\mathrm{imp},j} = -\sum_{i=1}^{\nu-1} N^i x_{2j,0} \delta_{\tau_j}^{(i-1)}.$$

Furthermore, we have $\mathcal{I}_{2,\tau_j}(w) = \sum_{i=1}^{\nu-1} \delta_{\tau_j}^{(i-1)} N^i w$. Thus, $\tilde{x}_{2,\mathrm{imp},j} = \mathcal{I}_{2,\tau_j}(w)$ if and only if $-Nx_{2j,0} = Nw$. Assuming that there exists an appropriate input u such that $-Nx_{2j,0} = Nw$, we get by using $x_{2,j,0} = \hat{x}_{2,j-1}(\tau_j)$ and $\hat{x}_{2,j-1}(\tau_j) = -\sum_{i=0}^{\nu-1} N^i B_2 \hat{u}_{j-1}^{(i)}(\tau_j)$ that

$$w \in \operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix} + \operatorname{Ker}(N).$$

We have the decomposition $w = \tilde{w} + \hat{w}$ with $\tilde{w} = \operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \cdots N^{\nu-1}B_2 \end{bmatrix}$, $\hat{w} \in \operatorname{Ker}(N)$ and $\tilde{w} = \sum_{j=0}^{\nu-1} N^j B_2 \tilde{w}_j$ for some $\tilde{w}_j \in \mathbb{R}^{n_\infty}$. Thus

$$\begin{aligned} \mathcal{I}_{2,\tau_j}(w) &= \sum_{i=1}^{\nu-1} \delta_{\tau_j}^{(i-1)} N^i (\tilde{w} + \hat{w}) = \sum_{i=1}^{\nu-1} \delta_{\tau_j}^{(i-1)} N^i \sum_{j=0}^{\nu-1} N^j B_2 \tilde{w}_j \\ &= \sum_{i=1}^{\nu-1} \sum_{j=0}^{\nu-1} \delta_{\tau_j}^{(i-1)} N^{i+j} B_2 \tilde{w}_j. \end{aligned}$$

which implies $\mathcal{I}_{2,\tau_j}(w) \in \operatorname{Im}\begin{bmatrix} NB_2 & N^2B_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix}$. On the other hand, if $\mathcal{I}_{2,\tau_j}(w) \in \operatorname{Im}\begin{bmatrix} NB_2 & N^2B_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix}$, then

$$N^i w \in \operatorname{Im}\begin{bmatrix} NB_2 & N^2B_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix} = N\operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \cdots & N^{\nu-2}B_2 \end{bmatrix}$$

for $i = 1, \ldots, \nu - 1$ and in particular

$$Nw \in N\operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \cdots & N^{\nu-2}B_2 \end{bmatrix} \iff w \in \operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix} + \operatorname{Ker}(N).$$

We can find an appropriate u by using Theorem 4.29. □

Definition 4.31. The system (4.8) is called *impulse controllable (*I-controllable) if for any initial condition $x(0)$, $\tau_j \in \mathbb{T}$ and $w \in \mathbb{R}^{n_\infty}$ there exists an admissible control function $u \in \mathcal{C}_p^\infty$ such that $x_{\mathrm{imp},j} = \mathcal{I}_{\tau_j}(w)$.

Theorem 4.32. *The following statements are equivalent.*

1. *The system* (4.8) *is I-controllable.*
2. *The fast subsystem* (4.9b) *is I-controllable.*
3. $Ker(N) + Im\begin{bmatrix} B_2 & NB_2 & \cdots & N^{\nu-1}B \end{bmatrix} = \mathbb{R}^{n_\infty}$.
4. $Im(N) = Im\begin{bmatrix} NB_2 & N^2B_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix}$.
5. $Im(N) + Im(B_2) + Ker(N) = \mathbb{R}^{n_\infty}$.
6. $rank\begin{bmatrix} E & 0 & 0 \\ A & E & B \end{bmatrix} = n + rank(E)$.

Proof.

1 $\iff$ **2** Clear since $x_{\text{imp}} = \begin{bmatrix} 0 \\ x_{\text{imp},2} \end{bmatrix}$.

2 $\iff$ **3** Follows from Theorem 4.30.

3 $\iff$ **4** Follows since

$$\begin{aligned} \operatorname{Im}\begin{bmatrix} NB_2 & N^2B_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix} &= N\operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix} \\ &= N\left(\operatorname{Im}\begin{bmatrix} B_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix} + \operatorname{Ker}(N)\right). \end{aligned}$$

4 $\implies$ **5** Follows from

$$\begin{aligned} \operatorname{Im}(N) + \operatorname{Im}(B_2) + \operatorname{Ker}(N) &= \operatorname{Im}\begin{bmatrix} NB_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix} + \operatorname{Im}(B_2) + \operatorname{Ker}(N) \\ &= \operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix} + \operatorname{Ker}(N) = \mathbb{R}^{n_\infty}, \end{aligned}$$

where the last equality follows from 3.

5 $\implies$ **3** It holds that $\operatorname{Im}(N^{i+1}) + \operatorname{Im}(N^iB_2) = \operatorname{Im}(N^i)$ for $i = 1, \ldots, \nu-1$. This implies

$$\begin{aligned} \mathbb{R}^{n_\infty} &= \operatorname{Ker}(N) + \operatorname{Im}(B_2) + \operatorname{Im}(NB_2) + \ldots + \operatorname{Im}(N^{\nu-1}B_2) + \operatorname{Im}(N^\nu) \\ &= \operatorname{Ker}(N) + \operatorname{Im}\begin{bmatrix} B_2 & NB_2 & \cdots & N^{\nu-1}B_2 \end{bmatrix}. \end{aligned}$$

4 $\iff$ **6** Consider the Kalman decomposition [5] of the matrix pair (N, B_2), i. e. there exists a nonsingular matrix V such that

$$(N, B_2) \sim (V^{-1}NV, V^{-1}B_2) = \left(\begin{bmatrix} N_{11} & N_{12} \\ 0 & N_{22} \end{bmatrix}, \begin{bmatrix} B_{21} \\ 0 \end{bmatrix}\right) \begin{matrix} \tilde{n}_1 \\ \tilde{n}_2 \end{matrix}$$

with (N_{11}, B_{21}) is C-controllable, i. e. $\operatorname{Im}\begin{bmatrix} B_{21} & N_{11}B_{21} & \cdots & N_{11}^{\nu-1}B_{21} \end{bmatrix} = \mathbb{R}^{\tilde{n}_1}$. Thus,

$$\begin{aligned} &\operatorname{Im}(N) = \operatorname{Im}\left(\begin{bmatrix} N_{11} & N_{12} \\ 0 & N_{22} \end{bmatrix}\right) \overset{(4)}{=} \operatorname{Im}\begin{bmatrix} NB_2 & N^2B_2 & \cdots N^{\nu-1}B_2 \end{bmatrix} = \operatorname{Im}\begin{bmatrix} N_{11} \\ 0 \end{bmatrix} \\ &\iff \tilde{n}_2 = 0 \text{ or } N_{22} = 0. \end{aligned}$$

Moreover, it holds that

$$\begin{aligned}
\operatorname{rank}\left(\begin{bmatrix} E & 0 & 0 \\ A & E & B \end{bmatrix}\right) &= \operatorname{rank}\left(\begin{bmatrix} WET & 0 & 0 \\ WAT & WET & WB \end{bmatrix}\right) \\
&= \operatorname{rank}\left(\begin{bmatrix} I_{n_f} & 0 & 0 & 0 & 0 \\ 0 & N & 0 & 0 & 0 \\ J & 0 & I_{n_f} & 0 & B_1 \\ 0 & I_{n_\infty} & 0 & N & B_2 \end{bmatrix}\right) \\
&= 2n_f + \operatorname{rank}\left(\begin{bmatrix} N & 0 & 0 \\ I_{n_\infty} & N & B_2 \end{bmatrix}\right) \\
&= 2n_f + \operatorname{rank}\left(\begin{bmatrix} N_{11} & N_{12} & 0 & 0 & 0 \\ 0 & N_{22} & 0 & 0 & 0 \\ I_{\tilde{n}_1} & 0 & N_{11} & N_{12} & B_{21} \\ 0 & I_{\tilde{n}_2} & 0 & N_{22} & 0 \end{bmatrix}\right).
\end{aligned}$$

We know that $\operatorname{rank}\begin{bmatrix} N_{11} & B_{21} \end{bmatrix} = \tilde{n}_1$. Hence, we have

$$\begin{aligned}
\operatorname{rank}\left(\begin{bmatrix} E & 0 & 0 \\ A & E & B \end{bmatrix}\right) &= 2n_f + \tilde{n}_1 + \tilde{n}_2 + \operatorname{rank}\left(\begin{bmatrix} N_{11} & -N_{12}N_{22} \\ 0 & -N_{22}^2 \end{bmatrix}\right) \\
&= n + n_f + \operatorname{rank}\left(\begin{bmatrix} N_{11} & N_{12}N_{22} \\ 0 & N_{22}^2 \end{bmatrix}\right).
\end{aligned}$$

Thus $\operatorname{rank}\left(\begin{bmatrix} E & 0 & 0 \\ A & E & B \end{bmatrix}\right) = n + \operatorname{rank}(E) = n + n_f + \operatorname{rank}\left(\begin{bmatrix} N_{11} & N_{12} \\ 0 & N_{22} \end{bmatrix}\right)$ if and only if

$$\operatorname{rank}\left(\begin{bmatrix} N_{11} & N_{12} \\ 0 & N_{22} \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} N_{11} & N_{12}N_{22} \\ 0 & N_{22}^2 \end{bmatrix}\right).$$

Since N_{22} is nilpotent, this consequently holds if and only if $\tilde{n}_2 = 0$ or $N_{22} = 0$.

□

Example 4.33. *Consider again the descriptor system from Example 4.13, given by*

$$\begin{aligned}
\dot{x}_1 &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \\
0 &= x_2 + \begin{bmatrix} -1 \\ 0 \end{bmatrix} u.
\end{aligned}$$

We have already seen that the system is R-controllable (Example 4.16) but not C-controllable (Example 4.13). We have

$$\begin{aligned}
Im(N) + Ker(N) + Im(B_2) &= Im\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + Ker\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + Im\begin{bmatrix} -1 \\ 0 \end{bmatrix} \\
&= \{0\} + \mathbb{R}^2 + Im\begin{bmatrix} -1 \\ 0 \end{bmatrix} = \mathbb{R}^2.
\end{aligned}$$

This implies that the system is I-controllable. Alternatively, we have

$$rank\left(\begin{bmatrix} E & 0 & 0 \\ A & E & B \end{bmatrix}\right) = rank\left(\begin{bmatrix} I_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ * & 0 & I_2 & 0 & e_2 \\ 0 & I_2 & 0 & 0 & -e_1 \end{bmatrix}\right) = 6 = n + rank(E).$$

Theorem 4.34. *The following statements are equivalent.*

1. *The system* (4.8) *is I-controllable.*
2. $rank\left(\begin{bmatrix} E & AS_\infty & B \end{bmatrix}\right) = n$ *where* S_∞ *is a matrix with* $Im(S_\infty) = Ker(N)$.
3. *There exists* $F \in \mathbb{R}^{m,n}$ *such that* $(E, A+BF)$ *is regular and* $\nu = ind(E, A+BF) \leq 1$.
4. $rank\left(\begin{bmatrix} N & K_\infty & B_2 \end{bmatrix}\right) = n_\infty$ *where* $Im(K_\infty) = Ker(N)$.
5. $rank\left(\begin{bmatrix} N & 0 & 0 \\ I_{n_\infty} & N & B_2 \end{bmatrix}\right) = n_\infty + rank(N)$.

Proof.

1 $\iff$ **2** As in the proof of the Theorem on feedback regularization (Theorem 3.1) there exist nonsingular matrices P and Q such that

$$PEQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}, \qquad PAQ = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad PB = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix}$$

and $\operatorname{rank}(E) = r$, $Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$. In particular, $Q_2 = S_\infty$. Thus, we have

$$\operatorname{rank}\left(\begin{bmatrix} E & AS_\infty & B \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} I_r & 0 & A_{12} & \tilde{B}_1 \\ 0 & 0 & A_{22} & \tilde{B}_2 \end{bmatrix}\right) = r + \operatorname{rank}\left(\begin{bmatrix} A_{22} & \tilde{B}_2 \end{bmatrix}\right).$$

Analogously,

$$\operatorname{rank}\left(\begin{bmatrix} E & 0 & 0 \\ A & E & B \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} I_r & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ A_{11} & A_{12} & I_r & 0 & \tilde{B}_1 \\ A_{21} & A_{22} & 0 & 0 & \tilde{B}_2 \end{bmatrix}\right) = 2r + \operatorname{rank}\left(\begin{bmatrix} A_{22} & \tilde{B}_2 \end{bmatrix}\right).$$

Thus, we have $\operatorname{rank}\left(\begin{bmatrix} E & AS_\infty & B \end{bmatrix}\right) = n$ if and only if $\operatorname{rank}\left(\begin{bmatrix} A_{22} & \tilde{B}_2 \end{bmatrix}\right) = n - r$ if and only if $\operatorname{rank}\left(\begin{bmatrix} E & 0 & 0 \\ A & E & B \end{bmatrix}\right) = n + r$.

2 $\iff$ **3** Clear (see Section 3.1).

2 $\iff$ **4** Let S_∞ such that $ES_\infty = 0$, then $WETT^{-1}S_\infty = \begin{bmatrix} I_{n_f} & 0 \\ 0 & N \end{bmatrix} T^{-1}S_\infty = 0$ with $T^{-1}S_\infty = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$. In particular, we have $S_1 = 0$ and $NS_2 = 0$, which implies that $S_2 = K_\infty$. The rest follows immediately as before.

1 $\iff$ **5** Follows from the previous theorem.

□

Definition 4.35. A system (4.8) is called *strongly controllable* (S-controllable) if it is R-controllable and I-controllable, i. e. if $\operatorname{rank}\begin{bmatrix}\lambda E - A & B\end{bmatrix} = n$ for all finite $\lambda \in \mathbb{C}$ and $\operatorname{rank}\begin{bmatrix}E & AS_\infty & B\end{bmatrix} = n$, where $\operatorname{Im}(S_\infty) = \ker(E)$.

The relationship between the various controllability concepts can be presented as

$$\begin{array}{rcll} & \Longrightarrow & \text{R-controllability} & \\ \text{C-controllability} & \Longrightarrow & \text{S-controllability} & \Longrightarrow \quad \text{R- and I-controllability} \\ & \Longrightarrow & \text{I-controllability} & \end{array}$$

Remark 4.36.

1. The ability to cancel all impulses in the system response by choosing a suitable state feedback control such that the resulting closed-loop system is regular and of index $\nu \leq 1$ is often used as definition for I-controllability. With this regard, linear time-varying and nonlinear descriptor systems can be handled as in chapter 3.
2. C-controllability and R-controllability for linear time-varying descriptor systems can be treated via appropriate staircase forms (see chapter 5).
3. Nonlinear descriptor systems are usually handled by local linearization.

4.2 Observability

Definition 4.37. The descriptor system (1.1) is called *completely observable (C-observable)* if the initial condition $x(0)$ can be uniquely determined by $u(t)$ and $y(t)$ for $0 \leq t < \infty$.

This means that for a C-controllable system the state $x(t)$ can be uniquely determined from u and y by observing the initial condition and constructing the system response at any time $t \geq 0$.

Definition 4.38 (Alternative definition)**.** The system (1.1) is C-observable if the zero output $y(t) \equiv 0$ with $u(t) \equiv 0$ implies that the system has only the trivial solution $x(t) \equiv 0$.

Definition 4.39. The system (1.1) is called *observable within the reachable set (R-observable)* if any state in the reachable set can be uniquely determined by $y(t)$ and $u(t)$ for $t \geq 0$.

Thus, we need an appropriate projection to variables that are associated with the dynamical part of the system. For general nonlinear systems this is difficult to obtain.

Remark 4.40.

1. R-observability is sometimes also called *finite dynamics observability*.
2. While C-observability reflects the reconstruction ability of the whole state $x(t)$ from measured output together with the control input, R-observability characterized the ability to reconstruct only the reachable states. Thus, C-observability implies R-observability.

Definition 4.41. A descriptor system (1.1) is called *impulse observable (I-observable)* if the impulse behavior in the state response $x(t)$ can be uniquely determined from the impulse behavior of the output and jump behavior in the input.

Theorem 4.42. *Consider a regular linear descriptor system of the form* (4.1).

1. *Let* $u(t) \equiv 0$. *Then* $y(t) \equiv 0$ *for all* $t \geq 0$ *if and only if*

$$\tilde{x}_0 \in Ker \begin{bmatrix} C_1 \\ C_1 J \\ \vdots \\ C_1 J^{n_f - 1} \end{bmatrix} \oplus Ker \begin{bmatrix} C_2 \\ C_2 N \\ \vdots \\ C_2 N^{\nu - 1} \end{bmatrix}, \qquad \tilde{x} = T^{-1} x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ as in (4.2).}$$

2. *The slow subsystem* (4.2a) *is C-observable if and only if* $rank \begin{bmatrix} \lambda E - A \\ C \end{bmatrix} = n$ *for all finite* $\lambda \in \mathbb{C}$.
3. *The following statements are equivalent:*
   a) *The fast subsystem* (4.2b) *is C-observable.*
   b) $rank \begin{bmatrix} C_2 \\ C_2 N \\ \vdots \\ C_2 N^{\nu - 1} \end{bmatrix} = n_\infty$.
   c) $Ker \begin{bmatrix} N \\ C_2 \end{bmatrix} = \{0\}$.
   d) $rank \begin{bmatrix} N \\ C_2 \end{bmatrix} = n_\infty$.
   e) $rank \begin{bmatrix} E \\ C \end{bmatrix} = n$.
   f) *For any two nonsingular matrices* P_1, Q_1 *satisfying*

$$P_1 E Q_1 = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \qquad \text{let} \quad C P_1 = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 \end{bmatrix}.$$

   Then $\tilde{C}_2$ *is of full column rank,* $rank(\tilde{C}_2) = n - rank(E)$.
4. *The following statements are equivalent:*
   a) *The system* (4.1) *is C-observable.*
   b) *The slow and fast subsystem* (4.2a) *and* (4.2b) *are both C-observable.*
   c) $rank \begin{bmatrix} \lambda E - A \\ C \end{bmatrix} = n$ *for all* $\lambda \in \mathbb{C}$ *and* $rank \begin{bmatrix} E \\ C \end{bmatrix} = n$.

d) rank $\begin{bmatrix} \alpha E - \beta A \\ C \end{bmatrix} = n$ *for all* $(\alpha, \beta) \in \mathbb{C}^2 \backslash \{(0,0)\}$.

Proof.

1. For $u(t) \equiv 0$ the state response of (4.2) is given by $x_1(t) = e^{Jt}x_1(0)$ and $x_2(t) = -\sum_{i=1}^{\nu-1} N^i x_2(0)\delta_0^{(i-1)}$ and $y(t) = y_1(t) + y_2(t) = C_1x_1(t) + C_2x_2(t)$. Thus

$$\begin{aligned} y(t) \equiv 0 &\iff C_1x_1(t) + C_2x_2(t) = 0 \quad \text{for all } t \geq 0 \\ &\iff C_1x_1(t) = 0 \qquad \text{and} \qquad C_2x_2(t) = 0 \quad \text{for all } t \geq 0, \end{aligned}$$

   where the last iff follows from the decomposition in the smooth and impulsive part. If $y_1(t) = C_1x_1(t) = C_1e^{Jt}x_1(0) \equiv 0$ for all $t \geq 0$, then also $y_1^{(i)}(t) \equiv 0$ for all $i = 0, \ldots, n_f - 1$ and all $t \geq 0$. In particular, we have $y_1^{(i)}(0) = 0$ and hence

$$\begin{bmatrix} C_1 \\ C_1J \\ \vdots \\ C_1J^{n_f-1} \end{bmatrix} x_1(0) = 0 \implies x_1(0) \in \operatorname{Ker} \begin{bmatrix} C_1 \\ C_1J \\ \vdots \\ C_1J^{n_f-1} \end{bmatrix}.$$

   Moreover,

$$\begin{aligned} y_2(t) = C_2x_2(t) = -\sum_{i=1}^{\nu-1} C_2N^i x_2(0)\delta_0^{(i-1)} &\equiv 0 \\ \iff C_2N^i x_2(0) = 0 \qquad &\text{for } i = 0, \ldots, \nu - 1 \\ \iff x_2(0) \in \operatorname{Ker} &\begin{bmatrix} C_2 \\ C_2N \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix}. \end{aligned}$$

   Thus, $\tilde{x}(0) = \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} \in \operatorname{Ker} \begin{bmatrix} C_1 \\ C_1J \\ \vdots \\ C_1J^{n_f-1} \end{bmatrix} \oplus \operatorname{Ker} \begin{bmatrix} C_2 \\ C_2N \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix}$.

2. The slow subsystem is a standard LTI system. Thus, (4.2a) is C-observable if and only if (J, C_1) is C-observable if and only if $\operatorname{rank} \begin{bmatrix} \lambda I - J \\ C_1 \end{bmatrix} = n_f$ for all $\lambda \in \mathbb{C}$. Furthermore

$$\begin{aligned} \operatorname{rank} \begin{bmatrix} \lambda E - A \\ C \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \lambda WET - WAT \\ CT \end{bmatrix} &= \operatorname{rank} \begin{bmatrix} \lambda I - J & 0 \\ 0 & \lambda N - I \\ C_1 & C_2 \end{bmatrix} \\ &= n_\infty + \operatorname{rank} \begin{bmatrix} \lambda I - J \\ C_1 \end{bmatrix}. \end{aligned}$$

3. By definition C-observability of the fast subsystem means that for $y_2(t) \equiv 0$ for $t \geq 0$ with $u(t) \equiv 0$ it follows that $x_2(0) = 0$. By 1 this is equivalent to $\operatorname{Ker} \begin{bmatrix} C_2 \\ C_2N \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix} = \{0\}$ and hence 3a $\iff$ 3b. Furthermore,

$$\operatorname{rank} \begin{bmatrix} C_2 \\ C_2N \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix} = n_\infty \iff \operatorname{rank} \begin{bmatrix} C_2^T & N^TC_2^T & \cdots & (N^T)^{\nu-1}C_2 \end{bmatrix} = n_\infty$$

$$\iff N^T\dot{\xi}_2 = \xi_2 + C^Tu \qquad \text{is C-controllable (Theorem 4.12).}$$

The above system is called the *dual system* of (4.2b). By Theorem 4.12 we have that 3b-3f are equivalent.

4. Follows from 1 and similar as in 3 by Theorem 4.12.

□

Example 4.43. *Consider again the descriptor system from Example 4.13 with additional output equation, given by*

$$\begin{aligned} \dot{x}_1 &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \\ 0 &= x_2 + \begin{bmatrix} -1 \\ 0 \end{bmatrix} u \\ y &= \begin{bmatrix} 1 & 0 \end{bmatrix} x_1. \end{aligned}$$

Since

$$\operatorname{rank} \begin{bmatrix} C_1 \\ C_1J \end{bmatrix} = \operatorname{rank} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = 2 = n_f \qquad \text{and} \qquad \operatorname{rank} \begin{bmatrix} C_2 \\ C_2N \end{bmatrix} = \operatorname{rank} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0 < n_\infty,$$

we find that the slow subsystem is C-observable while the fast subsystem is not.

Theorem 4.44. *Consider a regular linear system of the form* (4.1)*. Then the system* (4.1) *is R-observable if and only if the slow subsystem* (4.2a) *is C-observable, i. e.*

$$\operatorname{rank} \begin{bmatrix} \lambda E - A \\ C \end{bmatrix} = n \qquad \text{for all finite } \lambda \in \mathbb{C}.$$

Proof. Any reachable state $x(t) = T \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$ has the form

$$\begin{aligned} x_1(t) &= e^{Jt}x_1(0) + \int_0^t e^{J(t-s)}B_1u(s)ds, \\ x_2(t) &= -\sum_{i=0}^{\nu-1} N^i B_2 u^{(i)}(t), \\ y(t) &= y_1(t) + y_2(t) = C_1x_1(t) + C_2x_2(t), \end{aligned}$$

i. e. $x_2(t)$ is uniquely determined by $u(t)$ and $y_1(t) = C_1x_1(t) = y(t) - C_2x_2(t)$ is uniquely determined by $y(t)$ and $u(t)$. Thus, a reachable state $x(t)$ can be reconstructed from $y(t)$ and $u(t)$ if and only if $x_1(t)$ can be uniquely determined by $y_1(t)$ and $u(t)$, i. e. the slow subsystem (4.2a) is C-observable. □

Corollary 4.45. *The system* (4.1) *is C-observable if and only if the system is R-observable and rank* $\begin{bmatrix} E \\ C \end{bmatrix} = n$.

To prove the dual properties for I-observability we use the following lemma.

Lemma 4.46. *Consider a regular system in WCF* (4.2). *Then, the impulsive part of the output at* $\tau_j \in \mathbb{T}$ *is* $y_{imp,j} \equiv 0$ *for all* $j \in \mathbb{Z}$ *for* $u(t) \equiv 0$ *if and only if*

$$x_{2j,0} \in Ker \begin{bmatrix} C_2N \\ C_2N^2 \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix}, \qquad x_{2j,0} = \hat{x}_{2,j-1}(\tau_j), \tau_j \in \mathbb{T}.$$

Proof. Since $y_{\mathrm{imp},j} = Cx_{\mathrm{imp},j} = C_2x_{2,\mathrm{imp},j}$ we have $y_{\mathrm{imp},j} = -\sum_{i=1}^{\nu-1} C_2N^ix_{2j,0}\delta_{\tau_j}^{(i)}$ for all $\tau_j \in \mathbb{T}$. Thus, $y_{\mathrm{imp},j} \equiv 0$ if and only if $C_2N^ix_{2j,0} = 0$ for $i = 1, \ldots, \nu-1$ if and only if $x_{2j,0} \in \mathrm{Ker} \begin{bmatrix} C_2N \\ C_2N^2 \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix}$. □

Theorem 4.47. *Consider system* (4.1). *Then, the following statements are equivalent.*

1. *The system* (4.1) *is I-observable.*
2. *The fast subsystem* (4.2b) *is I-observable.*
3. $Ker \begin{bmatrix} C_2 \\ C_2N \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix} \cap Im(N) = \{0\}$.

4. $Ker \begin{bmatrix} C_2N \\ \vdots \\ C_2N\nu - 1 \end{bmatrix} = Ker(N)$.
5. $Ker(N) \cap Ker(C_2) \cap Im(N) = \{0\}$.
6. *Let*

$$\left(\begin{bmatrix} N_{11}^\top & N_{21}^\top \\ 0 & N_{22}^\top \end{bmatrix}, \begin{bmatrix} C_{21}^\top \\ 0 \end{bmatrix} \right) \begin{matrix} \tilde{n}_1 \\ \tilde{n}_2 \end{matrix}$$

be the Kalman decomposition [5] of $(N^\top, C_2^\top)$*, where* (N_{11}, C_{21}) *is C-observable. Then either* $\tilde{n}_2 = 0$ *or* $N_{22} = 0$ *and* $rank(N_{11}) = rank \begin{bmatrix} N_{11} \\ N_{21} \end{bmatrix}$.

7. $rank \begin{bmatrix} E & A \\ 0 & E \\ 0 & C \end{bmatrix} = n + rank(E)$.

Proof. Assume that $u \in \mathcal{C}_p^\infty$. From $x_{\text{imp}} = \begin{bmatrix} 0 \\ x_{2,\text{imp}} \end{bmatrix}$ and

$$y_{\text{imp}} = C_1 x_{1,\text{imp}} + C_2 x_{2,\text{imp}} = C_2 x_{2,\text{imp}} = y_{2,\text{imp}}.$$

Hence, we know 1 $\iff$ 2.

2 $\iff$ **4** For $u(t) \equiv 0$ we have $x_{2,\text{imp}} = -\sum_{i=1}^{\nu-1} N^i x_{2j,0} \delta_{\tau_j}^{(i-1)} \equiv 0 \iff Nx_{2j,0} = 0$. Thus from Lemma 4.46 we get 2 $\iff \text{Ker}(N) = \text{Ker} \begin{bmatrix} C_2N \\ C_2N^2 \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix} \iff$ 4.

3 $\iff$ **4** Assume that $\text{Ker} \begin{bmatrix} C_2N \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix} = \text{Ker}(N)$. Then for any

$$w \in \text{Ker} \begin{bmatrix} C_2 \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix} \cap \text{Im}(N),$$

there exists β such that

$$w = N\beta \in \text{Ker} \begin{bmatrix} C_2 \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix} \implies \beta \in \text{Ker} \begin{bmatrix} C_2N \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix} = \text{Ker}(N).$$

Conversely, assume that

$$\operatorname{Ker}\begin{bmatrix} C_2 \\ C_2N \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix} \cap \operatorname{Im}(N) = \{0\}.$$

Then, for any $w \in \operatorname{Ker}\begin{bmatrix} C_2N \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix}$ it holds that $Nw \in \operatorname{Ker}\begin{bmatrix} C_2 \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix} \subseteq \operatorname{Ker}(N)$. Since $\operatorname{Ker}(N) \subseteq \operatorname{Ker}\begin{bmatrix} C_2N \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix}$ the equality follows.

3 $\iff$ **5** Similar as before.

Assume that $\operatorname{Ker}(N) = \operatorname{Ker}\begin{bmatrix} CN_2 \\ \vdots \\ C_2N^{\nu-1} \end{bmatrix}$. Since $\operatorname{Ker}(N) + \operatorname{Im}(N^\top) = \mathbb{R}^{n_\infty}$ we have that $\operatorname{Im}(N^\top) = \operatorname{Im}\begin{bmatrix} N^\top C_2^\top & \cdots & \left(N^\top\right)^{\nu-1} C_2^\top \end{bmatrix}$. Then the equivalence of 4,5,6 and 7 follows from Theorem 4.32 (I-controllability). □

Definition 4.48. A descriptor system (4.1) is called *strongly observable* (S-observable) if it is R-observable and I-observable.

Note that controllability and observability are dual concepts, i. e. the following results hold:

Theorem 4.49 (Duality Principle)**.** *Consider a linear descriptor system of the form* (4.1)*. The the following holds:*

1. *The system* (4.1) *is C-controllable if and only if the dual system given by*

$$\begin{aligned} E^\top \dot{\xi} &= A^\top \xi + C^\top u \\ y &= B^\top \xi \end{aligned} \tag{4.10}$$

   is C-observable.
2. *The system* (4.1) *is R-controllable if and only if the dual system* (4.10) *is R-observable.*
3. *The system* (4.1) *is I-controllable if and only if the dual system* (4.10) *is I-observable.*

Proof. Follows from the previous discussion. □

The relationship between the different observability concepts are as follows:

$$\begin{array}{rcll} & \Longrightarrow & \text{R-observability} & \\ \text{C-observability} & \Longrightarrow & \text{S-observability} & \Longrightarrow \quad \text{R- and I-observability} \\ & \Longrightarrow & \text{I-observability} & \end{array}$$

Corollary 4.50. *The following statements are equivalent:*

1. *The system* (4.8) *is I-observable.*
2. $\mathrm{rank} \begin{bmatrix} E \\ T_\infty^\top A \\ C \end{bmatrix} = n$, *where* $Im(T_\infty) = Ker(E^\top)$.
3. $\mathrm{rank} \begin{bmatrix} N \\ K_\infty \\ C_2 \end{bmatrix} = n_\infty$, *where* $Im(K_\infty) = Ker(N^\top)$.

Remark 4.51. For the existence of a feedback $F \in \mathbb{R}^{m,p}$ such that $(E, A+BFC)$ is regular and of index $\nu \leq 1$ we need I-observability and I-controllability.

CHAPTER

5

STAIRCASE FORMS AND SYSTEM PROPERTIES

The results presented in the previous chapters that are based on the Weierstraß canonical form (WCF) are useful from a theoretical point of view. However, it is well known that the numerical computation of the (WCF) in finite precision is ill-conditioned and small perturbations can radically change the kind and number of blocks in the (WCF).

A better numerical way are staircase algorithms that use a sequence of rank decisions and transformations with orthogonal matrices to transform the system into a suitable condensed form.

We consider linear descriptor systems of the form

$$\begin{aligned} E\dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{5.1}$$

with $E, A \in \mathbb{R}^{l,n}$, $B \in \mathbb{R}^{l,m}$ and $C \in \mathbb{R}^{p,n}$.

Lemma 5.1. *There exist orthogonal matrices $P \in \mathbb{R}^{l,l}$ and $Q \in \mathbb{R}^{n,n}$ such that*

$$PEQ = \begin{bmatrix} E_{11} & 0 & E_{13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} r \\ s \\ l-r-s \end{matrix}, \qquad PAQ = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{bmatrix}, \qquad PB = \begin{bmatrix} B_1 \\ B_2 \\ 0 \end{bmatrix},$$
$$CQ = \begin{bmatrix} C_1 & C_2 & C_2 \end{bmatrix}$$

with $r = rank(E)$ and $s = rank(B_2)$.

Proof. By row compression there exists an orthogonal matrix $P \in \mathbb{R}^{l,l}$ such that

$$P\begin{bmatrix} E & B & A \end{bmatrix} = \left[\begin{array}{ccc|c|ccc} \tilde{E}_{11} & \tilde{E}_{12} & \tilde{E}_{13} & \tilde{B}_1 & \tilde{A}_{11} & \tilde{A}_{12} & \tilde{A}_{23} \\ 0 & 0 & 0 & \tilde{B}_2 & \tilde{A}_{21} & \tilde{A}_{22} & \tilde{A}_{23} \\ 0 & 0 & 0 & 0 & \tilde{A}_{31} & \tilde{A}_{32} & \tilde{A}_{33} \end{array}\right] \begin{matrix} r \\ s \\ l-r-s \end{matrix}$$

with $r = \operatorname{rank}(E)$ and $s = \operatorname{rank}(\tilde{B}_2)$ (e.g. by using SVD or QR decomposition). Then, we consider the matrix

$$\begin{bmatrix} \tilde{A}_{33} & \tilde{A}_{31} & \tilde{A}_{32} \\ \tilde{E}_{13} & \tilde{E}_{11} & \tilde{E}_{12} \end{bmatrix}.$$

There exists an orthogonal matrix $\tilde{Q} \in \mathbb{R}^{n,n}$ such that

$$\begin{bmatrix} \tilde{A}_{33} & \tilde{A}_{31} & \tilde{A}_{32} \\ \tilde{E}_{13} & \tilde{E}_{11} & \tilde{E}_{12} \end{bmatrix} \tilde{Q} = \begin{bmatrix} A_{33} & 0 & 0 \\ E_{13} & E_{11} & 0 \end{bmatrix} \begin{bmatrix} q = l-s-r \\ r \end{bmatrix}$$

(via column compression). Setting

$$Q = \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \\ I & 0 & 0 \end{bmatrix} \tilde{Q} \begin{bmatrix} 0 & 0 & I \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix}$$

it follows that PEQ, PAQ, PB are in the desired form. □

Theorem 5.2. *Consider (E, A, B, C) with $E, A \in \mathbb{R}^{n,n}$, $B \in \mathbb{R}^{n,m}$, $C \in \mathbb{R}^{p,n}$ (i. e. $l = n$). Then there exist orthogonal matrices P and Q such that*

$$\begin{aligned} PEQ &= \begin{bmatrix} E_{11} & 0 & E_{13} \\ 0 & 0 & E_{23} \\ 0 & 0 & E_{33} \end{bmatrix} \begin{matrix} t_1 \\ t_2, \\ t_3 \end{matrix} \qquad PAQ = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{bmatrix}, \qquad PB = \begin{bmatrix} B_1 \\ B_2 \\ 0 \end{bmatrix}, \\ CQ &= \begin{bmatrix} C_1 & C_2 & C_2 \end{bmatrix}, \end{aligned} \tag{5.2}$$

where $t_2 = n - t_1 - t_3$ and

1. *$\operatorname{rank}(E_{11}) = t_1$,*
2. *$\operatorname{rank}(B_2) = t_2$,*
3. *A_{33} is a block upper triangular matrix with square diagonal blocks and*
4. *E_{33} is a block upper triangular matrix with zero diagonal blocks and the same block decomposition as A_{33}.*

Proof. We inductively apply Lemma 5.1 starting with

$$\begin{aligned} P^{(1)}EQ^{(1)} &= \begin{bmatrix} E_{11}^{(1)} & 0 & E_{13}^{(1)} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} r \\ s, \\ q \end{matrix} \quad P^{(1)}AQ^{(1)} = \begin{bmatrix} A_{11}^{(1)} & A_{12}^{(1)} & A_{13}^{(1)} \\ A_{21}^{(1)} & A_{22}^{(1)} & A_{23}^{(1)} \\ 0 & 0 & A_{33}^{(1)} \end{bmatrix}, \quad P^{(1)}B = \begin{bmatrix} B_1^{(1)} \\ B_2^{(1)} \\ 0 \end{bmatrix}, \\ CQ^{(1)} &= \begin{bmatrix} C_1^{(1)} & C_2^{(1)} & C_2^{(1)} \end{bmatrix}. \end{aligned}$$

If $\operatorname{rank}(E_{11}^{(1)}) = r$ we have the desired form. Otherwise, we apply Lemma 5.1 for

$$\tilde{E} = \begin{bmatrix} E_{11}^{(1)} & 0 \\ 0 & 0 \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} A_{11}^{(1)} & A_{12}^{(1)} \\ A_{21}^{(1)} & A_{22}^{(1)} \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} B_1^{(1)} \\ B_2^{(1)} \end{bmatrix}, \quad \tilde{C} = \begin{bmatrix} C_1^{(1)} & C_2^{(1)} \end{bmatrix}$$

and obtain $\tilde{P}$ and $\tilde{Q}$. Setting

$$P^{(2)} = \begin{bmatrix} \tilde{P} & 0 \\ 0 & I \end{bmatrix} P^{(1)} \qquad \text{and} \qquad Q^{(2)} = Q^{(1)} \begin{bmatrix} \tilde{Q} & 0 \\ 0 & I \end{bmatrix}$$

we get

$$P^{(2)}EQ^{(2)} = \begin{bmatrix} E_{11}^{(2)} & 0 & E_{13}^{(2)} & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \end{bmatrix}, \qquad P^{(2)}AQ^{(2)} = \begin{bmatrix} A_{11}^{(2)} & A_{12}^{(2)} & A_{13}^{(2)} & A_{14}^{(2)} \\ A_{21}^{(2)} & A_{22}^{(2)} & A_{23}^{(2)} & A_{24}^{(2)} \\ 0 & 0 & A_{33}^{(2)} & A_{34}^{(2)} \\ 0 & 0 & 0 & A_{44}^{(2)} \end{bmatrix}.$$

We can continue inductively until $E_{11}^{(k)}$ has full rank. In every step the size of the blocks $E_{11}^{(k)}$ is decreased at least by 1, thus we obtain the desired form after a maximum of n steps. □

As a consequence of Theorem 5.2 we can separate the parts of the system (5.1) that are not I-controllable.

Corollary 5.3. *Consider the system* (5.1) *with matrices* E, A, B *transformed as in* (5.2) *(Theorem 5.2). Then the subsystem consisting of*

$$\begin{bmatrix} E_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u$$

is I-controllable.

Proof. With $S_\infty = \begin{bmatrix} 0 & I \end{bmatrix}^\top$ we have

$$\operatorname{rank}\left(\begin{bmatrix} E & AS_\infty & B \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} E_{11} & 0 & A_{12} & B_1 \\ 0 & 0 & A_{22} & B_2 \end{bmatrix}\right) = t_1 + t_2$$

of full rank. □

If we want to use output feedback control we also have to consider the matrix C.

Theorem 5.4. *Consider (E, A, B, C) as in Theorem 5.2. Then there exist orthogonal matrices $\tilde{P}$ and $\tilde{Q}$ such that*

$$\tilde{P}E\tilde{Q} = \begin{bmatrix} \tilde{E}_{11} & 0 & 0 & \tilde{E}_{14} \\ 0 & 0 & 0 & \tilde{E}_{24} \\ \tilde{E}_{31} & \tilde{E}_{23} & \tilde{E}_{33} & \tilde{E}_{34} \\ 0 & 0 & 0 & \tilde{E}_{44} \end{bmatrix} \begin{matrix} \tilde{t}_1 \\ \tilde{t}_2 \\ \tilde{t}_3 \\ \tilde{t}_4 \end{matrix}, \qquad \tilde{P}A\tilde{Q} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & 0 & \tilde{A}_{14} \\ \tilde{A}_{21} & \tilde{A}_{22} & 0 & \tilde{A}_{24} \\ \tilde{A}_{31} & \tilde{A}_{32} & \tilde{A}_{33} & \tilde{A}_{34} \\ 0 & 0 & 0 & \tilde{A}_{44} \end{bmatrix}$$
$$\tilde{P}B = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \\ \tilde{B}_3 \\ 0 \end{bmatrix} \begin{matrix} \tilde{t}_1 \\ \tilde{t}_2 \\ \tilde{t}_3 \\ \tilde{t}_4 \end{matrix}, \qquad C\tilde{Q} = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 & 0 & \tilde{C}_4 \end{bmatrix} \tag{5.3}$$

with

1. *$\operatorname{rank}(\tilde{E}_{11}) = \tilde{t}_1$,*
2. *$\operatorname{rank}(\tilde{C}_2) = \tilde{t}_2$,*
3. *$\tilde{A}_{33}$ and $\tilde{A}_{44}^\top$ are block lower triangular matrices with square diagonal blocks,*
4. *$\tilde{E}_{33}$ and $\tilde{E}_{44}^\top$ are block lower triangular matrices with zero diagonal blocks and the same dimensions as the diagonal blocks of $\tilde{A}_{33}$ and $\tilde{A}_{44}^\top$,*
5. *the subsystem consisting of*

$$\begin{bmatrix} \tilde{E}_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix} u$$
$$y = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

   is I-controllable and I-observable and
6. *the subsystem consisting of*

$$\begin{bmatrix} \tilde{E}_{11} & 0 & 0 \\ 0 & 0 & 0 \\ \tilde{E}_{31} & \tilde{E}_{32} & \tilde{E}_{33} \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & 0 \\ \tilde{A}_{21} & \tilde{A}_{22} & 0 \\ \tilde{A}_{31} & \tilde{A}_{32} & \tilde{A}_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \\ \tilde{B}_3 \end{bmatrix} u$$
$$y = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

   is I-controllable.

Proof. At first we determine P_1 and Q_1 such that P_1EQ_1, P_1AQ_1, P_1B and CQ_1 are in the form (5.2) of Theorem 5.2. Then, we consider the subsystem consisting of

$$\hat{E} = \begin{bmatrix} E_{11} & 0 \\ 0 & 0 \end{bmatrix}^\top, \qquad \hat{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^\top, \qquad \hat{B} = \begin{bmatrix} C_1 & C_2 \end{bmatrix}^\top$$

and apply Theorem 5.2 for $(\hat{E}, \hat{A}, \hat{B})$, i.e. we determine $\hat{P}$ and $\hat{Q}$ such that $\hat{P}\hat{E}\hat{Q}$, $\hat{P}\hat{A}\hat{Q}$ and $\hat{P}\hat{B}$ are in the form (5.2). Setting

$$P_2 = \begin{bmatrix} \hat{Q}^\top & 0 \\ 0 & I_{\tilde{t}_3} \end{bmatrix}, \qquad Q_2 = \begin{bmatrix} \hat{P}^\top & 0 \\ 0 & I_{\tilde{t}_3} \end{bmatrix}$$

and $\tilde{P} = P_2 P_1$, $\tilde{Q} = Q_1 Q_2$ we get the desired form (5.3). The properties 1-4 follow directly from Theorem 5.2. Properties 5 and 6 follow from Corollary 5.3 and the duality principle (Theorem 4.49). □

Based on the previous results we can decide if we can regularize the system by feedback control

Theorem 5.5. *Consider the system* (5.1) *with matrices* (E, A, B, C) *transformed as in* (5.2)*. Then there exists a matrix* $F \in \mathbb{R}^{m,n}$ *such that the pair* $(E, A + BF)$ *is regular if and only if* A_{33} *is nonsingular.*

Proof. Let F be partitioned as $F = \begin{bmatrix} F_1 & F_2 & F_3 \end{bmatrix}$. Then we have

$$\begin{aligned}
\det(\lambda E - (A + BF)) &= \det\left(\lambda \begin{bmatrix} E_{11} & 0 & E_{13} \\ 0 & 0 & E_{23} \\ 0 & 0 & E_{33} \end{bmatrix} - \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{bmatrix} - \begin{bmatrix} B_1 \\ B_2 \\ 0 \end{bmatrix} \begin{bmatrix} F_1 & F_2 & F_3 \end{bmatrix} \right) \\
&= \det\left(\left[\begin{array}{cc|c} \lambda E_{11} - A_{11} - B_1 F_1 & -A_{12} - B_1 F_2 & \lambda E_{13} - A_{13} - B_1 F_3 \\ -A_{21} - B_2 F_1 & -A_{22} - B_2 F_2 & \lambda E_{23} - A_{23} - B_2 F_3 \\ \hline 0 & 0 & \lambda E_{33} - A_{33} \end{array}\right] \right) \\
&= \det(\lambda E_{33} - A_{33}) \det\left(\lambda \begin{bmatrix} E_{11} & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} A_{11} + B_1 F_1 & A_{12} + B_1 F_2 \\ A_{21} + B_2 F_1 & A_{22} + B_2 F_2 \end{bmatrix} \right).
\end{aligned}$$

If A_{33} is singular, we have

$$\det(\lambda E_{33} - A_{33}) = \det\left(\lambda \begin{bmatrix} 0 & * & \dots & * \\ & \ddots & \ddots & \vdots \\ & & \ddots & * \\ & & & 0 \end{bmatrix} - \begin{bmatrix} * & \dots & \dots & * \\ & \ddots & & \vdots \\ & & \ddots & \vdots \\ & & & * \end{bmatrix} \right) = 0$$

for all $\lambda \in \mathbb{C}$. Otherwise, if A_{33} is nonsingular, we have $\det(\lambda E_{33} - A_{33}) \neq 0$ for all $\lambda \in \mathbb{C}$. Since B_2 has full rank there exists F_2 such that $A_{22} + B_2 F_2$ is nonsingular. Let $F = \begin{bmatrix} 0 & F_2 & 0 \end{bmatrix}$, then

$$\begin{aligned}
\det(\lambda E - (A + BF)) &= \det(\lambda E_{33} - A_{33}) \det\left(\lambda \begin{bmatrix} E_{11} & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} A_{11} + B_1 F_1 & A_{12} + B_1 F_2 \\ A_{21} + B_2 F_1 & A_{22} + B_2 F_2 \end{bmatrix} \right) \\
&= \det(\lambda E_{33} - A_{33}) \underbrace{\det(\lambda E_{11} - \tilde{A}_{11})}_{\not\equiv 0} \underbrace{\det(A_{22} + B_2 F_2)}_{\neq 0} \not\equiv 0.
\end{aligned}$$

□

Theorem 5.6. *Consider* (5.1) *with* (E, A, B, C) *given in the form* (5.2) *and* A_{33} *is nonsingular. Then there exists* $F \in \mathbb{R}^{m,n}$ *such that* $(E, A + BF)$ *is regular and*

$$ind(E, A + BF) = ind\left(\begin{bmatrix} 0 & E_{23} \\ 0 & E_{33} \end{bmatrix}\right).$$

Proof. Let $F = \begin{bmatrix} F_1 & F_2 & F_3 \end{bmatrix}$, then we have

$$(E, A + BF) = \left(\begin{bmatrix} E_{11} & 0 & E_{13} \\ 0 & 0 & E_{23} \\ 0 & 0 & E_{33} \end{bmatrix}, \begin{bmatrix} A_{11} + B_1F_1 & A_{12} + B_1F_2 & A_{13} + B_1F_3 \\ A_{21} + B_2F_1 & A_{22} + B_2F_2 & A_{23} + B_2F_3 \\ 0 & 0 & A_{33} \end{bmatrix}\right).$$

Since B_2 has full rank we can choose $F_1 \in \mathbb{R}^{m,t_1}$ and $F_2 \in \mathbb{R}^{m,t_2}$ such that $A_{21} + B_2F_1 = 0$ and $A_{22} + B_2F_2$ is nonsingular. With $F_3 = 0$ we get

$$(E, A + BF) = \left(\begin{bmatrix} E_{11} & 0 & E_{13} \\ 0 & 0 & E_{23} \\ 0 & 0 & E_{33} \end{bmatrix}, \begin{bmatrix} A_{11} + B_1F_1 & A_{12} + B_1F_2 & A_{13} \\ 0 & A_{22} + B_2F_2 & A_{23} \\ 0 & 0 & A_{33} \end{bmatrix}\right),$$

where $(E_{11}, A_{11} + B_1F_1)$ is regular and of index 0 since E_{11} is nonsingular. For the second block we have $A_{22} + B_2F_2$ and A_{33} nonsingular such that the index is given by

$$\operatorname{ind}(E, A + BF) = \operatorname{ind}\left(\begin{bmatrix} 0 & E_{23} \\ 0 & E_{33} \end{bmatrix}\right).$$

□

Remark 5.7.

1. The index $\operatorname{ind}\left(\begin{bmatrix} 0 & E_{23} \\ 0 & E_{33} \end{bmatrix}\right)$ is the minimal index we can reach by state feedback. If we want to obtain a closed-loop system of index 1 we need that $\begin{bmatrix} E_{23} \\ E_{33} \end{bmatrix} = 0$ (and A_{33} is nonsingular).
2. If $\begin{bmatrix} E_{23}^\top & E_{33}^\top \end{bmatrix}^\top \neq 0$ and A_{33} nonsingular, we have to consider a system of the form

$$\begin{bmatrix} E_{11} & 0 & E_{13} \\ 0 & 0 & E_{23} \\ 0 & 0 & E_{33} \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \\ 0 \end{bmatrix} u.$$

   From the last block equation we get $A_{33}^{-1}E_{33}\dot{x}_3 = x_3$ and $A_{33}^{-1}E_{33}$ is block upper triangular with zero blocks on the diagonal and therefore we have $x_3 \equiv 0$. This means that the part of the system that is not I-controllable, i. e. x_3, is not problematic since it vanishes identically (in particular is stable). The remaining system is I-controllable (see Corollary 5.3).
3. If A_{33} is not invertible the system cannot be regularized. In general this is assumed to be a modeling error.

Theorem 5.8. *Consider* (5.1) *with* (E, A, B, C) *given in the form* (5.2) *and* A_{33} *nonsingular and* $\begin{bmatrix} A_{22}^\top & C_2^\top \end{bmatrix}^\top$ *of full column rank. Then there exists* $F \in \mathbb{R}^{m,p}$ *such that* $(E, A+BFC)$ *is regular and*

$$ind(E, A + BFC) = ind\left(\begin{bmatrix} 0 & E_{23} \\ 0 & E_{33} \end{bmatrix}\right).$$

Proof. Since B_2 has full row rank and $\begin{bmatrix} A_{22}^\top & C_2^\top \end{bmatrix}^\top$ has full column rank, there exists F such that $A_{22} + B_2FC_2$ is nonsingular. Then the result follows in the same way as in the proof of Theorem 5.5 and Theorem 5.6. □

Remark 5.9. In the case of output feedback we can consider the system in the form (5.3) of Theorem 5.4 given by

$$\begin{bmatrix} \tilde{E}_{11} & 0 & 0 & \tilde{E}_{14} \\ 0 & 0 & 0 & \tilde{E}_{24} \\ \tilde{E}_{31} & \tilde{E}_{23} & \tilde{E}_{33} & \tilde{E}_{34} \\ 0 & 0 & 0 & \tilde{E}_{44} \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & 0 & \tilde{A}_{14} \\ \tilde{A}_{21} & \tilde{A}_{22} & 0 & \tilde{A}_{24} \\ \tilde{A}_{31} & \tilde{A}_{32} & \tilde{A}_{33} & \tilde{A}_{34} \\ 0 & 0 & 0 & \tilde{A}_{44} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \\ \tilde{B}_3 \\ 0 \end{bmatrix} u$$

$$y = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 & 0 & \tilde{C}_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

Similar as before the system can be regularized by output feedback if and only if $\tilde{A}_{33}$ and $\tilde{A}_{44}$ are nonsingular. In this case we get $x_4 \equiv 0$, i. e. the part that is not I-controllable vanishes identically. The part x_3 belonging to the third block equation is dangerous since there can occur derivatives of the input u that cannot be observed if the index is not 0 (i. e. $\tilde{t}_3 > 0$ and $\tilde{E}_{33} \neq 0$).

Justified by the previous discussion it is often assumed that the system (5.1) is I-controllable and I-observable. If this is not the case then the parts that are not I-controllable and not I-observable can be removed using the transformation into the form (5.2) or (5.3). However, under the presence of rounding errors these components might still cause trouble.

Theorem 5.10 (Staircase form for linear descriptor systems, [2])**.** *Let* $E, A \in \mathbb{R}^{n,n}$, $B \in \mathbb{R}^{n,m}$ *and* $C \in \mathbb{R}^{p,n}$. *Then there exist orthogonal matrices* $U, V \in \mathbb{R}^{n,n}$, $W \in \mathbb{R}^{m,m}$ *and*

$Y \in \mathbb{R}^{p,p}$ *such that*

$$U^\top EV = \begin{array}{c} \\ t_1 \\ n-t_1 \end{array} \begin{array}{c} \begin{array}{cc} t_1 & n-t_1 \end{array} \\ \begin{bmatrix} \Sigma_E & 0 \\ 0 & 0 \end{bmatrix} \end{array}, \quad U^\top AV = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{array} \begin{array}{c} \begin{array}{cccccc} t_1 & s_2 & t_5 & t_4 & t_3 & s_6 \end{array} \\ \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \\ A_{21} & A_{22} & A_{23} & A_{24} & 0 & 0 \\ A_{31} & A_{32} & A_{33} & A_{34} & \Sigma_{35} & 0 \\ A_{41} & A_{42} & A_{43} & \Sigma_{44} & 0 & 0 \\ A_{51} & 0 & \Sigma_{53} & 0 & 0 & 0 \\ A_{61} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array},$$

$$U^\top BW = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{array} \begin{array}{c} \begin{array}{ccc} k_1 & k_2 & k_3 \end{array} \\ \begin{bmatrix} B_{11} & B_{12} & 0 \\ B_{21} & 0 & 0 \\ B_{31} & 0 & 0 \\ 0 & 0 & \\ 0 & 0 & \\ 0 & 0 & \end{bmatrix} \end{array}, \quad Y^\top CV = \begin{array}{c} \\ l_1 \\ l_2 \\ l_3 \end{array} \begin{array}{c} \begin{array}{cccccc} t_1 & s_2 & t_5 & t_4 & t_3 & s_6 \end{array} \\ \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{21} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \tag{SCF}$$

and the matrices $\Sigma_E, \Sigma_{35}, \Sigma_{44}, \Sigma_{53}$ *are nonsingular diagonal,* B_{12} *has full column rank,* C_{21} *has full row rank and the matrices*

$$\begin{bmatrix} B_{21} \\ B_{31} \end{bmatrix} \in \mathbb{R}^{k_1,k_1} \qquad \text{and} \qquad \begin{bmatrix} C_{12} & C_{13} \end{bmatrix} \in \mathbb{R}^{l_1,l_1}$$

with $k_1 = t_2 + t_3$ *and* $l_1 = s_2 + t_5$ *are nonsingular.*

Before we prove this Theorem, we draw several conclusions.

Corollary 5.11. *Let* E, A *be in staircase form* (SCF)*. Then the following statements are equivalent:*

1. *The pair* (E, A) *is regular and* $\nu = ind(E, A) \leq 1$.
2. $s_6 = t_6 = 0$ *and* A_{22} *nonsingular.*
3. $rank \begin{bmatrix} E & AS_\infty \end{bmatrix} = n$, *where* $range(S_\infty) = Ker(E)$.
4. $rank \begin{bmatrix} E \\ T_\infty^\top A \end{bmatrix} = n$, *where* $range(T_\infty) = Ker(E^\top)$.

Proof. We have

$$\begin{bmatrix} E & AS_\infty \end{bmatrix} = \begin{bmatrix} \Sigma_E & 0 & \hat{A}_{12} \\ 0 & 0 & \hat{A}_{22} \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} E \\ T_\infty^\top A \end{bmatrix} = \begin{bmatrix} \Sigma_E & 0 \\ 0 & 0 \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix},$$

where

$$\hat{A}_{12} = \begin{bmatrix} A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \end{bmatrix}$$

$$\hat{A}_{22} = \begin{bmatrix} A_{22} & A_{23} & A_{24} & 0 & 0 \\ A_{32} & A_{33} & A_{34} & \Sigma_{35} & 0 \\ A_{42} & A_{43} & \Sigma_{44} & 0 & 0 \\ 0 & \Sigma_{53} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \hat{A}_{21} = \begin{bmatrix} A_{21} \\ A_{31} \\ A_{41} \\ A_{51} \\ A_{61} \end{bmatrix}.$$

Thus, the equivalences 2, 3 and 4 follows.

2 $\implies$ **1** If $s_6 = t_6 = 0$ and A_{22} nonsingular, then $\hat{A}_{22}$ is nonsingular. Thus,

$$(E, A) \sim \left(\begin{bmatrix} \Sigma_E & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} A_{11} - \hat{A}_{12}\hat{A}_{22}\hat{A}_{21} & 0 \\ 0 & I \end{bmatrix} \right),$$

which is nonsingular and of index $\nu \leq 1$.

1 $\implies$ **3** If (E, A) is regular and of index $\nu \leq 1$ it holds

$$(E, A) \sim \left(\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix} \right)$$

and hence $\operatorname{rank}\left(\begin{bmatrix} E & AS_\infty \end{bmatrix}\right) = n$.

□

Corollary 5.12. *Let (E, A, B, C) be in staircase form* (SCF)*. Then it holds:*

1. *The system* (5.1) *is I-controllable if and only if $t_6 = 0$.*
2. *The system is I-observable if and only if $s_6 = 0$.*
3. *$\operatorname{rank}\begin{bmatrix} E & B \end{bmatrix} = n$ if and only if $t_4 = t_5 = t_6 = 0$.*
4. *$\operatorname{rank}\begin{bmatrix} E \\ C \end{bmatrix} = n$ if and only if $t_4 = t_3 = s_6 = 0$.*
5. *The system* (5.1) *is C-controllable if and only if $t_4 = t_5 = t_6 = 0$ and the system* (5.1) *is R-controllable (i. e. $\operatorname{rank}\begin{bmatrix} \lambda E - A & B \end{bmatrix} = n$ for all $\lambda \in \mathbb{C}$).*
6. *The system* (5.1) *is C-observable if and only if $t_4 = t_3 = s_6 = 0$ and the system* (5.1) *is R-observable.*

Proof. Follows directly from the staircase form (SCF). □

Remark 5.13. We assume without loss of generality that $k_3 = l_3 = 0$ in the staircase form (SCF) since these parts have no influence on the system and can be omitted by defining a new input u or a new output y.

For the proof of Theorem 5.10 we can proceed using the Algorithm 1.

Algorithm 1 Staircase Form

Input: $E, A \in \mathbb{R}^{n,n}$, $B \in \mathbb{R}^{n,m}$, $C \in \mathbb{R}^{p,n}$
Output: orthogonal matrices $U, V \in \mathbb{R}^{n,n}$, $W \in \mathbb{R}^{m,m}$, $Y \in \mathbb{R}^{p,p}$ such that $U^\top EV$, $U^\top AV$, $U^\top BW$ and $Q^\top CV$ are in staircase form (SCF).

Set $U := I_n, V := I_n, W := I_m, Y := I_p$.
Step 1 Compute a SVD of E

$$E = U_E \begin{bmatrix} \Sigma_E & 0 \\ 0 & 0 \end{bmatrix} V_E^\top$$

with Σ_E of size $t_1 \times t_1$ nonsingular and diagonal. Update

$$E := U_E^\top E V_E = \begin{array}{c} \\ t_1 \\ n-t_1 \end{array} \begin{array}{c} \begin{array}{cc} t_1 & n-t_1 \end{array} \\ \begin{bmatrix} \Sigma_E & 0 \\ 0 & 0 \end{bmatrix} \end{array}, \quad A := U_E^\top A V_E = \begin{array}{c} \\ t_1 \\ n-t_1 \end{array} \begin{array}{c} \begin{array}{cc} t_1 & n-t_1 \end{array} \\ \begin{bmatrix} A_{11}^{(1)} & A_{12}^{(1)} \\ A_{21}^{(1)} & A_{22}^{(1)} \end{bmatrix} \end{array},$$

$$B := U_E^\top B = \begin{array}{c} \\ t_1 \\ n-t_1 \end{array} \begin{array}{c} \begin{array}{c} m \end{array} \\ \begin{bmatrix} B_1^{(1)} \\ B_2^{(1)} \end{bmatrix} \end{array}, \qquad C := C V_E = \begin{array}{c} \\ p \end{array} \begin{array}{c} \begin{array}{cc} t_1 & n-t_1 \end{array} \\ \begin{bmatrix} C_1^{(1)} & C_2^{(1)} \end{bmatrix} \end{array}$$

and set $U := UU_E, V := VV_E$.
Step 2 Compute SVDs of $B_2^{(1)}$ and $C_2^{(1)}$:

$$B_2^{(1)} = U_B \begin{bmatrix} \Sigma_B & 0 \\ 0 & 0 \end{bmatrix} V_B^\top, \qquad C_2^{(1)} = U_C \begin{bmatrix} \Sigma_C & 0 \\ 0 & 0 \end{bmatrix} V_C^\top$$

with Σ_B of size $k_1 \times k_1$, Σ_C of size $l_1 \times l_1$ nonsingular and diagonal. Define

$$U_1 = \begin{bmatrix} I_{t_1} & 0 \\ 0 & U_B^\top \end{bmatrix} \qquad \text{and} \qquad V_1 = \begin{bmatrix} I_{t_1} & 0 \\ 0 & V_C \end{bmatrix}$$

and perform the updates

$$E := U_1 E V_1 = \begin{array}{c} \\ t_1 \\ k_1 \\ n-t_1-k_1 \end{array} \begin{array}{c} \begin{array}{ccc} t_1 & l_1 & n-t_1-l_1 \end{array} \\ \begin{bmatrix} \Sigma_E & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array},$$

$$A := U_1 A V_1 = \begin{array}{c} \\ t_1 \\ k_1 \\ n-t_1-k_1 \end{array} \begin{array}{c} \begin{array}{ccc} t_1 & l_1 & n-t_1-l_1 \end{array} \\ \begin{bmatrix} A_{11}^{(2)} & A_{12}^{(2)} & A_{13}^{(2)} \\ A_{21}^{(2)} & A_{22}^{(2)} & A_{23}^{(2)} \\ A_{31}^{(2)} & A_{32}^{(2)} & A_{33}^{(2)} \end{bmatrix} \end{array},$$

$$B := U_1 B V_B = \begin{array}{c} \\ t_1 \\ k_1 \\ n-t_1-k_1 \end{array} \begin{array}{c} \begin{array}{cc} t_1 & m-t_1 \end{array} \\ \begin{bmatrix} B_{11}^{(2)} & B_{12}^{(2)} \\ \Sigma_B & 0 \\ 0 & 0 \end{bmatrix} \end{array},$$

$$C := U_C^\top C V_1 = \begin{array}{c} \\ l_1 \\ p-l_1 \end{array} \begin{array}{c} \begin{array}{ccc} t_1 & l_1 & n-t_1-l_1 \end{array} \\ \begin{bmatrix} C_{11}^{(2)} & \Sigma_C & 0 \\ C_{21}^{(2)} & 0 & 0 \end{bmatrix} \end{array}$$

and $U := UU_1, V := VV_1, Y := YU_C, W := WV_B$.

Step 3 Compute SVDs of $B_{12}^{(2)}$ and $C_{21}^{(2)}$:

$$B_{12}^{(2)} = U_{12} \begin{bmatrix} \Sigma_{12} & 0 \\ 0 & 0 \end{bmatrix} V_{12}^\top, \qquad C_{21}^{(2)} = U_{21} \begin{bmatrix} \Sigma_{21} & 0 \\ 0 & 0 \end{bmatrix} V_{21}^\top$$

with Σ_{12} of size $k_2 \times k_2$ and Σ_{21} of size $l_2 \times l_2$. Update

$$B := \begin{bmatrix} I_{k_1} & 0 \\ 0 & V_{12} \end{bmatrix} = \begin{array}{c} \\ t_1 \\ k_1 \\ n-t_1-k_1 \end{array} \begin{array}{c} \begin{array}{ccc} k_1 & k_2 & k_3 \end{array} \\ \begin{bmatrix} B_{11}^{(3)} & B_{12}^{(3)} & 0 \\ \Sigma_B & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array},$$

$$C := \begin{bmatrix} I_{l_1} & 0 \\ 0 & U_{21}^\top \end{bmatrix} C = \begin{array}{c} \\ l_1 \\ l_2 \\ l_3 \end{array} \begin{array}{c} \begin{array}{ccc} t_1 & l_1 & n-t_1-l_1 \end{array} \\ \begin{bmatrix} C_{11}^{(3)} & \Sigma_C & 0 \\ C_{21}^{(3)} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

Step 4 Compute SVD of $A_{33}^{(2)}$:

$$A_{33}^{(2)} = U_A \begin{bmatrix} \Sigma_{44} & 0 \\ 0 & 0 \end{bmatrix} V_A^\top$$

with Σ_{44} of size $t_4 \times t_4$ nonsingular an diagonal. Set $n_{4,1} = n - t_1 - k_1 - t_4$,

$n_{4,2} = n - t_1 - l_1 - t_4$ and perform the following updates:

$$E := \begin{bmatrix} I_{t_1} & & \\ & I_{k_1} & \\ & & U_A^\top \end{bmatrix} E \begin{bmatrix} I_{t_1} & & \\ & I_{l_1} & \\ & & V_A \end{bmatrix} = \begin{array}{c} \\ t_1 \\ k_1 \\ t_4 \\ n_{4,2} \end{array} \begin{array}{c} \begin{array}{cccc} t_1 & l_1 & t_4 & n_{4,1} \end{array} \\ \begin{bmatrix} \Sigma_E & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$A := \begin{bmatrix} I_{t_1} & & \\ & I_{k_1} & \\ & & U_A^\top \end{bmatrix} A \begin{bmatrix} I_{t_1} & & \\ & I_{l_1} & \\ & & V_A \end{bmatrix} = \begin{bmatrix} A_{11}^{(33)} & A_{12}^{(33)} & A_{13}^{(33)} & A_{14}^{(33)} \\ A_{21}^{(33)} & A_{22}^{(33)} & A_{23}^{(33)} & A_{24}^{(33)} \\ A_{31}^{(33)} & A_{32}^{(33)} & \Sigma_{44} & 0 \\ A_{41}^{(33)} & A_{42}^{(33)} & 0 & 0 \end{bmatrix}$$

$$B := \begin{bmatrix} I_{t_1} & & \\ & I_{k_1} & \\ & & U_A^\top \end{bmatrix} B = \begin{array}{c} \\ t_1 \\ k_1 \\ t_4 \\ n_{4,2} \end{array} \begin{array}{c} \begin{array}{ccc} k_1 & k_2 & k_3 \end{array} \\ \begin{bmatrix} B_{11}^{(3)} & B_{12}^{(3)} & 0 \\ \Sigma_B & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$C := C \begin{bmatrix} I_{t_1} & & \\ & I_{l_1} & \\ & & V_A \end{bmatrix} = \begin{array}{c} \\ l_1 \\ l_2 \\ l_3 \end{array} \begin{array}{c} \begin{array}{cccc} t_1 & l_1 & t_4 & n_{4,1} \end{array} \\ \begin{bmatrix} C_{11}^{(3)} & \Sigma_C & 0 & 0 \\ C_{21}^{(3)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$U := U \begin{bmatrix} I_{t_1} & & \\ & I_{k_1} & \\ & & U_A \end{bmatrix}, \qquad V := V \begin{bmatrix} I_{t_1} & & \\ & I_{l_1} & \\ & & V_A \end{bmatrix}$$

Step 5 Compute a permuted SVD of $A_{42}^{(3)}$ and $A_{24}^{(3)}$:

$$A_{42}^{(3)} = U_{42} \begin{bmatrix} 0 & \Sigma_{53} \\ 0 & 0 \end{bmatrix} V_{42}^\top, \qquad A_{24}^{(3)} = U_{24} \begin{bmatrix} 0 & 0 \\ \Sigma_{35} & 0 \end{bmatrix} V_{24}^\top$$

with Σ_{53} of size $t_5 \times t_5$ and Σ_{35} of size $t_3 \times t_3$ nonsingular and diagonal. We set $t_2 = k_1 - t_3$, $t_6 = n - \sum_{i=1}^{5} t_i$ and

$$V_5 = \begin{bmatrix} I_{t_1} & & & \\ & V_{42} & & \\ & & I_{t_4} & \\ & & & V_{24} \end{bmatrix} \qquad \text{and} \qquad U_5 = \begin{bmatrix} I_{t_1} & & & \\ & U_{24} & & \\ & & I_{t_4} & \\ & & & U_{42} \end{bmatrix}$$

perform the updates

$$E := U_5^\top E V_5 = \begin{bmatrix} \Sigma_E & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A := U_5^\top A V_5 = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \\ A_{21} & A_{22} & A_{23} & A_{24} & 0 & 0 \\ A_{31} & A_{32} & A_{33} & A_{34} & \Sigma_{35} & 0 \\ A_{41} & A_{42} & A_{43} & \Sigma_{44} & 0 & 0 \\ A_{51} & 0 & \Sigma_{53} & 0 & 0 & 0 \\ A_{61} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B := U_5^\top B = \begin{bmatrix} B_{11} & B_{12} & 0 \\ B_{21} & 0 & 0 \\ B_{31} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{where } \begin{bmatrix} B_{21} \\ B_{31} \end{bmatrix} \text{ is nonsinguar}$$

$$C := C V_5 = \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{21} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and $V := V V_5$, $U := U U_5$.

Remark 5.14.

1. The condensed form (SCF) can be computed by 8 SVDs.
2. Since only transformations with orthogonal matrices are used the algorithm is numerically backward stable.
3. Still, numerical rank decisions are critical since small perturbations can change the numerical rank drastically.
4. If one is only interested in state feedback control, then we can set $C = I$ and neglect the transformations that operate on C.

With the help of the condensed form (SCF) we can now construct feedback controls. In the following, we consider proportional and derivative output feedbacks of the form

$$u = Fy - G\dot{y} + v$$

for system (5.1). Then the closed-loop system has the form

$$(E + BGC)\dot{x} = (A + BFC)x + Bv$$

Theorem 5.15. *Let (E, A, B, C) be given in the form* (SCF).

1. *If the system is I-controllable and I-observable, i. e.*

$$\operatorname{rank}\begin{bmatrix} E & AS_\infty & B \end{bmatrix} = \operatorname{rank}\begin{bmatrix} E \\ T_\infty^\top A \\ C \end{bmatrix} = n,$$

then for all $s \in \mathbb{N}$ with $0 \le s \le t_2 = s_2$ there exist matrices $F, G \in \mathbb{R}^{m,p}$ such that $(E + BGC, A + BFC)$ is regular and $\nu = ind(E + BGC, A + BFC) \le 1$ and $rank(E + BGC) = t_1 + s$.
   a) If $s = t_2$, then this is achieved by derivative feedback alone with $F = 0$.
   b) If $s = 0$, then this is achieved by proportional feedback alone with $G = 0$.
2. *If there exists $F \in \mathbb{R}^{m,p}$ such that $(E, A+BFC)$ is regular and $\nu = ind(E, A+BFC) \le 1$, then*

$$\operatorname{rank}\begin{bmatrix} E & AS_\infty & B \end{bmatrix} = \operatorname{rank}\begin{bmatrix} E \\ T_\infty^\top \\ C \end{bmatrix} = n,$$

i. e. the system is I-controllable and I-observable.

Remark 5.16. For the case of proportional and derivative state feedback we can set $C = I$.

Proof.

1. We can assume that (E, A, B, C) are in the form (SCF) with $k_3 = l_3 = 0$ and $t_6 = s_6 = 0$ (Corollary 5.12) as well as $t_2 = s_2$. Let

$$G = \begin{array}{c} \\ k_1 \\ m - k_1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} l_1 & p - l_1 \end{array} \\ \begin{bmatrix} G_{11} & 0 \\ 0 & 0 \end{bmatrix} \end{array} \in \mathbb{R}^{m,p}, \qquad F = \begin{array}{c} \\ k_1 \\ m - k_1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} l_1 & p - l_1 \end{array} \\ \begin{bmatrix} F_{11} & 0 \\ 0 & 0 \end{bmatrix} \end{array} \in \mathbb{R}^{m,p}$$

with

$$G_{11} = \begin{bmatrix} B_{21} \\ B_{31} \end{bmatrix}^{-1} \begin{bmatrix} I_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} C_{12} & C_{13} \end{bmatrix}^{-1} \in \mathbb{R}^{k_1, l_1}$$

$$F_{11} = \begin{bmatrix} B_{21} \\ B_{31} \end{bmatrix}^{-1} \begin{array}{c} \\ s \\ t_2 - s \\ k_1 - t_2 \end{array}\!\!\begin{array}{c} \begin{array}{ccc} s & t_2 - s & l_1 - t_2 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \begin{bmatrix} C_{12} & C_{13} \end{bmatrix}^{-1} \in \mathbb{R}^{k_1, l_1}$$

and $\phi = I_{t_2 - s} - \begin{bmatrix} 0 & I_{t_2-s} \end{bmatrix} A_{22} \begin{bmatrix} 0 \\ I_{t_2-s} \end{bmatrix} \in \mathbb{R}^{t_2-s, t_2-s}$. Then we have

$$BGC = \begin{array}{c} \\ t_1 \\ s \\ n - t_1 - s \end{array}\!\!\begin{array}{c} \begin{array}{ccc} t_1 & s & n - t_1 - s \end{array} \\ \begin{bmatrix} \Delta_{11} & \Delta_{12} & 0 \\ \Delta_{21} & \Delta_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

with

$$\begin{aligned}
\Delta_{11} &= B_{11}G_{11}C_{11} \in \mathbb{R}^{t_1,t_1} \\
\Delta_{12} &= B_{11}\begin{bmatrix} B_{21} \\ B_{31} \end{bmatrix}^{-1}\begin{bmatrix} I_s \\ 0 \end{bmatrix} \in \mathbb{R}^{t_1,s} \\
\Delta_{21} &= \begin{bmatrix} I_s & 0 \end{bmatrix}\begin{bmatrix} C_{12} & C_{13} \end{bmatrix}^{-1} \in \mathbb{R}^{s,t_1} \\
\Delta_{22} &= I_s
\end{aligned}$$

and

$$BFC = \begin{array}{c} \\ t_1 \\ s \\ t_2 - s \\ n - t_1 - t_2 \end{array}\begin{array}{c} \begin{array}{cccc} t_1 & s & t_2 - s & n - t_1 - t_2 \end{array} \\ \begin{bmatrix} \phi_{11} & 0 & \phi_{13} & 0 \\ 0 & 0 & 0 & 0 \\ \phi_{31} & 0 & \phi_{33} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \in \mathbb{R}^{n,n}$$

with

$$\begin{aligned}
\phi_{11} &= B_{11}F_{11}C_{11} \in \mathbb{R}^{t_1,t_1} \\
\begin{bmatrix} 0 & C_{13} & 0 \end{bmatrix} &= B_{11}\begin{bmatrix} B_{21} \\ B_{31} \end{bmatrix}^{-1}\begin{bmatrix} 0 & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{t_1,n-t_1} \\
\begin{bmatrix} 0 \\ C_{31} \\ 0 \end{bmatrix} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} C_{12} & C_{13} \end{bmatrix}^{-1} C_{11} \in \mathbb{R}^{n-t_1,t_1} \\
\phi_{33} &= \phi.
\end{aligned}$$

It follows that

$$E - BGC = \begin{bmatrix} \Sigma_E + \Delta_{11} & \Delta_{12} & 0 \\ \Delta_{21} & I_s & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} I_{t_1} & \Delta_{12} & 0 \\ 0 & I_s & 0 \\ 0 & 0 & I \end{bmatrix}\begin{bmatrix} \Sigma_E + \Delta_{11} - \Delta_{12}\Delta_{21} & 0 & 0 \\ \Delta_{21} & I_s & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and $\operatorname{rank}(E + BGC) = s + \operatorname{rank}(\Sigma_E + \Delta_{11} - \Delta_{12}\Delta_{21})$. Furthermore, we have

$$\Delta_{11} - \Delta_{12}\Delta_{21} = B_{11}G_{11}C_{11} - B_{11}\underbrace{\begin{bmatrix} B_{21} \\ B_{31} \end{bmatrix}^{-1}\begin{bmatrix} I_s \\ 0 \end{bmatrix}\begin{bmatrix} I_s & 0 \end{bmatrix}\begin{bmatrix} C_{12} & C_{13} \end{bmatrix}^{-1}}_{=G_{11}} C_{11} = 0$$

and thus $\operatorname{rank}(E + BGC) = s + \operatorname{rank}(\Sigma_E) = s + t_1$. Due to the form of BFC we get

$$A + BFC = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} \\ A_{21} & A_{22} & A_{23} & A_{24} & 0 \\ A_{31} & A_{32} & A_{33} & A_{34} & \Sigma_{35} \\ A_{41} & A_{42} & A_{43} & \Sigma_{44} & 0 \\ A_{51} & 0 & \Sigma_{53} & 0 & 0 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \begin{bmatrix} 0 & \phi_{13} \end{bmatrix} & 0 & 0 & 0 \\ \begin{bmatrix} 0 \\ \phi_{31} \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & \phi_{33} \end{bmatrix} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & A_{13} & A_{14} & A_{15} \\ \tilde{A}_{21} & \tilde{A}_{22} & A_{23} & A_{24} & 0 \\ A_{31} & A_{32} & A_{33} & A_{34} & \Sigma_{35} \\ A_{41} & A_{42} & A_{43} & \Sigma_{44} & 0 \\ A_{51} & 0 & \Sigma_{53} & 0 & 0 \end{bmatrix}$$

with

$$\tilde{A}_{22} = A_{22} + \begin{bmatrix} 0 & 0 \\ 0 & \phi_{33} \end{bmatrix} = \begin{bmatrix} A_{22,1} & A_{22,2} \\ A_{22,3} & A_{22,4} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & I_{t_2-s} - A_{22,4} \end{bmatrix} = \begin{bmatrix} A_{22,1} & A_{22,2} \\ A_{22,3} & I_{t_2-s}. \end{bmatrix}$$

In particular,

$$\begin{array}{c} \\ t_2 - s \\ t_3 \\ t_4 \\ t_5 \end{array} \begin{array}{c} \begin{array}{cccc} t_2 - s & t_5 & t_4 & t_3 \end{array} \\ \begin{bmatrix} I_{t_2-s} & \tilde{A}_{23} & \tilde{A}_{24} & 0 \\ \tilde{A}_{32} & A_{33} & A_{34} & \Sigma_{35} \\ \tilde{A}_{42} & A_{43} & \Sigma_{44} & 0 \\ 0 & \Sigma_{53} & 0 & 0 \end{bmatrix} \end{array}$$

as the lower right $(n - t_1 - s) \times (n - t_1 - s)$ principle of $A + BFC$ is nonsingular. Thus, we have that $(E + BGC, A + BFC)$ is regular and of index $\nu \leq 1$. Moreover, it is clear from the construction that $F = 0$ if $s = t_2$ and $G = 0$ if $s = 0$.

2. If only proportional feedback is used then the matrices S_∞ and T_∞ are not changed by the feedback. Thus

$$\operatorname{rank} \begin{bmatrix} E & AS_\infty & B \end{bmatrix} = \operatorname{rank} \begin{bmatrix} E & (A + BFC)S_\infty & B \end{bmatrix} = \operatorname{rank} \begin{bmatrix} E \\ T_\infty^\top A \\ C \end{bmatrix}$$

$$= \operatorname{rank} \begin{bmatrix} E \\ T_\infty^\top (A + BFC) \\ C \end{bmatrix} = n$$

(see Theorem 3.1).

□

Remark 5.17.

1. If we use derivative feedback, then the existence of a feedback matrix G such that $(E + BGC, A)$ is regular and of index ≤ 1 is not sufficient for the system to be I-controllable and I-observable, since left and right nullspaces of E may change under derivative feedback.

2. To compute the regularizing feedback we need F_{11}, G_{11} as defined in the previous proof. Due to the construction of the form (SCF) the matrices $\begin{bmatrix} B_{21} \\ B_{31} \end{bmatrix}$ and $\begin{bmatrix} C_{12} & C_{13} \end{bmatrix}$ can be kept in factorized form as a product of an orthogonal matrix and a diagonal matrix. Thus for the computation of F_{11}, G_{11} or $E + BGC$, $A + BFC$ we only have to invert two diagonal matrices.

Corollary 5.18. *Let (E, A, B, C) be given in the form* (SCF)*. If*

$$\operatorname{rank} \begin{bmatrix} \lambda E - A \\ C \end{bmatrix} = n \qquad \text{for all } \lambda \in \mathbb{C}$$

and

$$\operatorname{rank} \begin{bmatrix} E & AS_\infty & B \end{bmatrix} = \operatorname{rank} \begin{bmatrix} E \\ T_\infty^\top A \\ C \end{bmatrix} = n$$

(i. e. the system is S-controllable and S-observable), then there exist $F, G \in \mathbb{R}^{m,p}$ and a feedback control $u = Fy - G\dot{y} + v$ such that the closed-loop system $(E + BGC, A + BFC)$ is S-controllable and S-observable with index ≤ 1 and $\operatorname{rank}(E + BGC) = t_1 + s$, where s is given such that $0 \leq s \leq t_2$.

Proof. From I-controllability and I-observability there follows the existence of F, G such that $(E + BGC, A + BFC)$ is regular and of index ≤ 1 and still I-controllable and I-observable. The other conditions are invariant under feedback, thus the closed-loop system is S-controllable and S-observable. □

Corollary 5.19. *Let (E, A, B, C) be given in the form* (SCF)*.*

1. *There exists G such that $E + BGC$ is nonsingular if and only if*

$$\operatorname{rank} \begin{bmatrix} E & B \end{bmatrix} = \operatorname{rank} \begin{bmatrix} E \\ C \end{bmatrix} = n.$$

2. *There exists $G \in \mathbb{R}^{m,p}$ and a control $u = -G\dot{y} + v$ such that the closed-loop system is C-controllable and C-observable with $\operatorname{rank}(E + BGC)$ if and only if*

$$\operatorname{rank} \begin{bmatrix} \alpha E - \beta A & B \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \alpha E - \beta A \\ C \end{bmatrix} = n \qquad \text{for all } (\alpha, \beta) \in \mathbb{C}^2 \backslash \{(0,0)\}.$$

Proof.

1. Assume first that there exists G such that $E + BGC$ is nonsingular. Then from the form (SCF) we get $t_3 = t_4 = t + 5 = t_6 = s_6 = 0$ and hence

$$\operatorname{rank} \begin{bmatrix} E \\ C \end{bmatrix} = \operatorname{rank} \begin{bmatrix} E & B \end{bmatrix} = n.$$

   The converse direction follows from Theorem 5.15 by choosing $s = t_2$.

2. Since the rank conditions are invariant under feedback, the claim follows.

□

Remark 5.20.

1. For linear systems with variable coefficients of the form

$$E(t)\dot{x}(t) = A(t)x(t) + B(t)u(t), \qquad x(t_0) = x_0$$

condensed/staircase forms have been derived in [3] using special kinds of global analytic singular value decompositions (ASVDs) [1]. For the existence of an ASVD we have to assume that $E(t), A(t)$ and $B(t)$ are analytic and $\operatorname{rank} E(t) = r$ for all $t \in \mathbb{I}$, then

$$E(t) = U(t) \begin{bmatrix} \Sigma(t) & 0 \\ 0 & 0 \end{bmatrix} V(t)^\top,$$

where $U(t), V(t)$ are analytic and pointwise orthogonal. Regularization by proportional or derivative feedback can then be handled in a similar manner.

2. Nonlinear systems can be treated by local linearization.

Remark 5.21. In principle we now have two possibilities to check some of the system properties: either we use the derivative array approach as in chapter 3 or the staircase form (SCF). The computation of the staircase form (SCF) is much more subtle numerically since the consecutive rank decisions of transformed matrices have to be made in a proper way. Also two-sided transformations are used that change the basis of the state space. The staircase form allows to check observability and controllability simultaneously but on the cost of changing the physical meaning of the state variables.

CHAPTER

6

OPTIMAL CONTROL PROBLEMS

We consider optimal control problems of the following form

$$\text{Minimize } \mathcal{J}(x,u) = \mathcal{M}(x(t_f)) + \int_{t_0}^{t_f} \mathcal{K}(t,x(t),u(t))dt \tag{6.1a}$$

$$\text{subject to } F(t,x,\dot{x},u) = 0, \quad x(t_0) = x_0 \tag{6.1b}$$

with $F \in \mathcal{C}(\mathbb{I}\times\mathbb{D}_x\times\mathbb{D}_{\dot{x}}\times\mathbb{D}_u, \mathbb{R}^l)$ sufficiently smooth, $\mathbb{I} = [t_0, t_f] \subseteq \mathbb{R}$, $\mathbb{D}_x, \mathbb{D}_{\dot{x}} \subseteq \mathbb{R}^n$, $\mathbb{D}_u \subseteq \mathbb{R}^m$ open sets, $x_0 \in \mathbb{D}_x$, $\mathcal{M} : \mathbb{D}_x \to \mathbb{R}$ and $\mathcal{K} : \mathbb{I} \times \mathbb{D}_x \times \mathbb{D}_u \to \mathbb{R}$.

BIBLIOGRAPHY

[1] A. Bunse-Gerstner, R. Byers, V. Mehrmann, and N. K. Nichols. Numerical computation of an analytic singular value decomposition of a matrix valued function. *Numer. Math.*, 60:1–40, 1991.

[2] A. Bunse-Gerstner, V. Mehrmann, and N. K. Nichols. Regularization of descriptor systems by output feedback. *IEEE Trans. Automat. Control*, 39:1742–1748, 1994.

[3] R. Byers, P. Kunkel, and V. Mehrmann. Regularization of linear descriptor systems with variable coefficients. *SIAM J. Cont.*, 35:117–133, 1997.

[4] J. D. Cobb. On the solutions of linear differential equations with singular coefficients. *J. Diff. Equations*, 46:310–323, 1982.

[5] R.E. Kalman. Canonical structure of linear dynamical systems. *Proc. Natl. Acad. Sci. U. S. A.*, 48(4):596–600, 1962.

[6] P. Kunkel and V. Mehrmann. *Differential-Algebraic Equations. Analysis and Numerical Solution*. European Mathematical Society, 2006.

INDEX