Technische Universität Chemnitz
Sonderforschungsbereich 393
Numerische Simulation auf massiv parallelen Rechnern

Peter Kunkel Volker Mehrmann

Analysis of over- and underdetermined nonlinear differential-algebraic systems with application to nonlinear control problems

Preprint SFB393/00-08

Preprint-Reihe des Chemnitzer SFB 393

SFB393/00-08 **February 2000**

Contents

Authors:

Peter Kunkel
Fachbereich Mathematik
Carl von Ossietzky Universität Oldenburg
Postfach 2503,
D-26111 Oldenburg,
Fed. Rep. Germany

Volker Mehrmann
Fakultät für Mathematik
Technische Universität Chemnitz
D-09107 Chemnitz
Fed. Rep. Germany

Analysis of over- and underdetermined nonlinear differential-algebraic systems with application to nonlinear control problems*

Abstract

We study over- and underdetermined systems of nonlinear differential-algebraic equations. Such equations arise in many applications in circuit and multibody system simulation, in particular when automatic model generation is used, or in the analysis and solution of control problems via a behaviour approach.

We give a general (local) existence and uniqueness theory and apply the results to nonlinear control problems. In particular, we study regularization by state or output feedback.

The theoretical analysis also leads immediately to numerical methods for the simulation as well as the construction of regularizing controls.

Keywords: nonlinear differential-algebraic equations, nonlinear control problems, solvability, model consistency, behaviour approach, strangeness index, regularization, feedback design

AMS(MOS) subject classification: 93C50, 65L05, 34H05, 93B10, 93B11, 93B40

1 Introduction

In this paper, we study nonlinear differential-algebraic systems of the form

$$F(t, x, \dot{x}) = 0, \tag{1}$$

with $F \in C(\mathbb{I} \times \mathbb{D}_x \times \mathbb{D}_{\dot{x}}, \mathbb{R}^m)$, $\mathbb{I} \subseteq \mathbb{R}$ (compact) interval, $\mathbb{D}_x, \mathbb{D}_{\dot{x}} \subseteq \mathbb{R}^n$ open. Such systems include in particular nonlinear control problems

$$F(t, \xi, u, \dot{\xi}) = 0, \tag{2}$$

$$y = G(t, \xi). \tag{3}$$

Here $\xi \in \mathbb{R}^{n_\xi}$ is the state, $u \in \mathbb{R}^{n_u}$ the input and $y \in \mathbb{R}^{n_y}$ the output of the system. Control systems of the form (2) can be rewritten in the form (1) via a behaviour approach that combines the vector functions ξ and u as

$$x = \begin{bmatrix} \xi \\ u \end{bmatrix} \in \mathbb{R}^n,$$

*Supported by DFG research grant Ku964/4-1: *Über- und unterbestimmte nichtlineare DAEs.*

with $n = n_\xi + n_u$, see [27, 29]. For the case of control problems that include the output equation (3), we set

$$x = \begin{bmatrix} \xi \\ u \\ y \end{bmatrix} \in \mathbb{R}^n$$

with $n = n_\xi + n_u + n_y$. We will discuss both cases.

Example 1 Throughout this paper we will illustrate the various results and problems by means of the simple nonlinear control system given by

$$F(t, \xi, u, \dot{\xi}) = \begin{bmatrix} \dot{\xi}_2 \\ \log \xi_2 + \sin u \end{bmatrix} = 0, \tag{4}$$

with $\xi = (\xi_1, \xi_2)^T$, $n_\xi = 2$ and $n_u = 1$. The corresponding behaviour system reads

$$F(t, x, \dot{x}) = \begin{bmatrix} \dot{x}_2 \\ \log x_2 + \sin x_3 \end{bmatrix} = 0, \tag{5}$$

with $n = 3$.

General models like (1) as well as the described control problems arise in mechanical multibody systems [14, 18, 19], electrical circuits [17] or mixed systems, where different models are coupled together [15]. In this general form they allow to model very complex dynamical systems with constraints, models that are automatically generated with redundant equations or combinations of models of different types, see, e.g., [15, 18]. Redundant equations typically occur when several submodels (modules) are linked together such that they form loops. We also get overdetermined systems when we include first integrals of the problem. For example, Hamiltonian systems of ordinary differential equations are known to conserve energy. Standard integration schemes typically cannot yield numerical approximations that keep the initial energy. Instead, the scheme produces or consumes energy during the integration. If conservation of energy is crucial, one can use so-called symplectic integration schemes (see, e. g., [20]), but one has to be very cautious with the stepsize selection. Combining the given system with the equation for the conservation of energy leads to an overdetermined system of differential-algebraic equations. Designing numerical methods for this type of problems that satisfy all inherent algebraic constraints is therefore an important topic.

To analyze general nonlinear differential-algebraic systems and also to design controls, we need to develop the mathematical theory as well as numerical methods that can be used for the analysis, design and simulation.

The theory and numerical solution methods for differential-algebraic equations have undergone major changes in the last 10 years, see [1, 8, 9, 16, 22, 23, 24, 26, 28, 30, 31]. The theory and also numerical techniques have also been partially extended to the study of linear control problems [6, 11, 12, 27, 32]. Except for the restriction to linear problems, a major drawback of the previous methods was the missing analysis for over- and underdetermined systems. It is the topic of this paper to extend the previous analysis of [22, 25, 26, 27] to the general nonlinear case.

The paper is organized as follows. After some preliminaries in Section 2, we give an analysis of the general problem in Section 3. We then apply these results to control problems in Section 4. In Section 5, we propose numerical algorithms for the integration of the arising problems. Finally, we give some conclusions in Section 6.

2 Preliminaries

Since the concepts for differential-algebraic equations (DAEs) have changed in recent years, we need to recall some of the terminology and some of the previous results.

Definition 1 Consider system (1). A function $x : \mathbb{I} \to \mathbb{R}^n$ is called a *solution* of (1) if $x \in C^1(\mathbb{I}, \mathbb{R}^n)$ and x satisfies (1) pointwise. It is called a *solution of the initial value problem* consisting of (1) and

$$x(t_0) = x_0, \tag{6}$$

if x is a solution of (1) and satisfies (6). An initial condition (6) is called *consistent* if the corresponding initial value problem has at least one solution.

In the control setting, for a given input function u the concept of solvability is described by Definition 1. Using the behaviour approach, it also covers the so-called *model consistency*, i. e., the existence of a solution ξ for some input function u. A more interesting question in the context of control problems is whether it is possible to choose a control such that the resulting problem is regular (in the sense of a DAE, see [26]). We will discuss this question at least locally, i. e., for a sufficiently small neighborhood of $t_0 \in \mathbb{I}$.

In order to analyze the properties of the system, like existence and uniqueness of solutions, in [26] for the square nonlinear case and in [27] for the rectangular linear case, hypotheses have been formulated which lead to an index concept, the so-called *strangeness-index* or *s-index* which generalizes the concept of differentiation index. In the following, we will formulate a generalization of these hypotheses and the strangeness-index for the general nonlinear nonsquare case. To do this, we assume for convenience that all functions are sufficiently smooth. As in [26], we introduce a nonlinear derivative array, see also [8, 10], of the form

$$F_\ell(t, x, \dot{x}, \ldots, x^{(\ell+1)}) = 0, \tag{7}$$

which stacks the original equation and all its derivatives up to level ℓ in one large system, i. e.,

$$F_\ell(t, x, \dot{x}, \ldots, x^{(\ell+1)}) = \begin{bmatrix} F(t, x, \dot{x}) \\ \frac{d}{dt} F(t, x, \dot{x}) \\ \vdots \\ \frac{d^\ell}{dt^\ell} F(t, x, \dot{x}) \end{bmatrix}. \tag{8}$$

Partial derivatives of F_ℓ with respect to selected variables p from $(t, x, \dot{x}, \ldots, x^{(\ell+1)})$ are denoted by $F_{\ell;p}$, e. g.,

$$F_{\ell;x} = \frac{\partial}{\partial x} F_\ell, \quad F_{\ell;\dot{x},\ldots,x^{(\ell+1)}} = [\, \frac{\partial}{\partial \dot{x}} F_\ell \ \ \ldots \ \ \frac{\partial}{\partial x^{(\ell+1)}} F_\ell \,].$$

A corresponding notation is used for partial derivatives of other functions.

3 General analysis

In order to analyze existence and uniqueness of solutions we need to study the solution set of the derivative array F_μ for some integer μ. We denote this set as

$$\mathbb{L}_\mu = \{ z_\mu \in \mathbb{I} \times \mathbb{R}^n \times \mathbb{R}^n \times \ldots \times \mathbb{R}^n \mid F_\mu(z_\mu) = 0 \}. \tag{9}$$

The following hypothesis extends Hypotheses 2.1 and 3.2 in [26], see also [24].

Hypothesis 1 *Consider the general system of nonlinear differential-algebraic equations (1). There exist integers μ, r, a, d, and v such that $\mathbb{L}_\mu$ is not empty, and the following properties hold:*

1. *The set $\mathbb{L}_\mu \subseteq \mathbb{R}^{(\mu+2)n+1}$ forms a manifold of dimension $(\mu+2)n+1-r$.*
2. *We have*
$$\operatorname{rank} F_{\mu;x,\dot{x},\ldots,x^{(\mu+1)}} = r \tag{10}$$
on $\mathbb{L}_\mu$.
3. *We have*
$$\operatorname{corank} F_{\mu;x,\dot{x},\ldots,x^{(\mu+1)}} - \operatorname{corank} F_{\mu-1;x,\dot{x},\ldots,x^{(\mu)}} = v \tag{11}$$
on $\mathbb{L}_\mu$. (The corank is the dimension of the corange and we use the convention that $\operatorname{corank} F_{-1;x} = 0$*.)*
4. *We have*
$$\operatorname{rank} F_{\mu;\dot{x},\ldots,x^{(\mu+1)}} = r - a \tag{12}$$
on $\mathbb{L}_\mu$ such that there are smooth full rank matrix functions Z_2 and T_2 defined on $\mathbb{L}_\mu$ of size $((\mu+1)m, a)$ and $(n, n-a)$, respectively, satisfying
$$Z_2^T F_{\mu;\dot{x},\ldots,x^{(\mu+1)}} = 0, \quad \operatorname{rank} Z_2^T F_{\mu;x} = a, \quad Z_2^T F_{\mu;x} T_2 = 0 \tag{13}$$
on $\mathbb{L}_\mu$.
5. *We have*
$$\operatorname{rank} F_{\dot{x}} T_2 = d = m - a - v \tag{14}$$
on $\mathbb{L}_\mu$.

For square systems without redundancies, i. e., $m = n$ and $v = 0$, Hypothesis 1 reduces to Hypothesis 3.2 in [26] and for linear time varying system to Hypothesis 2.7 in [24]. The difference to the assumptions in [10] are that we allow redundancies, underdeterminedness and that we do not require constant rank in a neighborhood of the solution in the whole space but only on a submanifold. Furthermore we need less smoothness of the function F. The latter observation is described in detail in [26]. As in [24, 26], we call the smallest possible μ the *strangeness-index* of (1). Systems with vanishing strangeness-index are called *strangeness-free*.

To derive the implications of Hypothesis 1 and to motivate the various assumptions, we proceed as follows.

Let $z_\mu^0 = (t_0, x_0, \dot{x}_0, \ldots, x_0^{(\mu+1)}) \in \mathbb{L}_\mu$. Observe that in this context $\dot{x}_0, \ldots, x_0^{(\mu+1)}$ denote algebraic variables in $\mathbb{R}^n$. Since $\mathbb{L}_\mu$ is a manifold of dimension $(\mu+2)n+1-r$, we can locally parametrize it by $(\mu+2)n+1-r$ parameters. These can be chosen from $(t, x, \dot{x}, \ldots, x^{(\mu+1)})$ in such a way that discarding the associated columns from

$$F_{\mu;t,x,\dot{x},\ldots,x^{(\mu+1)}}(t_0, x_0, \dot{x}_0, \ldots, x_0^{(\mu+1)})$$

does not lead to a rank drop. Because of part 2 of Hypothesis 1, $F_{\mu;x,\dot{x},\ldots,x^{(\mu+1)}}$ has already maximal rank. Hence, we can always choose t as a parameter.

Because of part 4 of Hypothesis 1, we can choose $n - a$ parameters out of x. Without restriction we can write x as (x_1, x_2, x_3) with $x_1 \in \mathbb{R}^d$, $x_2 \in \mathbb{R}^{n-a-d}$, $x_3 \in \mathbb{R}^a$, and choose (x_1, x_2) as further parameters. In particular, the matrix $Z_2^T F_{\mu;x_3}$ is then nonsingular. The remaining parameters $p \in \mathbb{R}^{(\mu+1)n+a-r}$ can be chosen out of $(\dot{x}, \ldots, x^{(\mu+1)})$.

Hence, Hypothesis 1 implies that there is a diffeomorphism φ defined on a neighborhood $\mathbb{U} \subseteq \mathbb{R}^{(\mu+2)n+1-r}$ of $(t_0, x_{10}, x_{20}, p_0)$ as part of z_μ^0 corresponding to the selected parameters (t, x_1, x_2, p) and a neighborhood $\mathbb{V} \subseteq \mathbb{R}^{(\mu+2)n+1}$ of z_μ^0 such that

$$\mathbb{L} \cap \mathbb{V} = \{\varphi(t, x_1, x_2, p) \mid (t, x_1, x_2, p) \in \mathbb{U}\}.$$

This includes that locally $F_\mu(z_\mu) = 0$ if and only if $z_\mu = \varphi(t, x_1, x_2, p)$ for some $(t, x_1, x_2, p) \in \mathbb{U}$. In particular, there are functions $\mathcal{G}$ (corresponding to x_3) and $\mathcal{H}$ (corresponding to $(\dot{x}, \ldots, x^{(\mu+1)})$) such that

$$F_\mu(t, x_1, x_2, \mathcal{G}(t, x_1, x_2, p), \mathcal{H}(t, x_1, x_2, p)) \equiv 0 \tag{15}$$

on $\mathbb{U}$.

Defining

$$\hat{F}_2 = Z_2^T F_\mu \tag{16}$$

on $\mathbb{U}$, where Z_2 is given according to Hypothesis 1, we have

$$\hat{F}_2(t, x_1, x_2, \mathcal{G}(t, x_1, x_2, p), \mathcal{H}(t, x_1, x_2, p)) \equiv 0 \tag{17}$$

on $\mathbb{U}$. Differentiation with respect to p yields (omitting arguments)

$$\begin{aligned} \tfrac{d}{dp}\hat{F}_2 &= (Z_{2;x_3}^T F_\mu + Z_2^T F_{\mu;x_3})\mathcal{G}_p + (Z_{2;\dot{x},\ldots,x^{(\mu+1)}}^T F_\mu + Z_2^T F_{\mu;\dot{x},\ldots,x^{(\mu+1)}})\mathcal{H}_p \\ &= Z_2^T F_{\mu;x_3}\mathcal{G}_p \equiv 0 \end{aligned}$$

on $\mathbb{U}$. By construction, the parameters x_3 were selected such that $Z_2^T F_{\mu;x_3}$ is nonsingular. Thus,

$$\mathcal{G}_p(t, x_1, x_2, p) \equiv 0$$

on $\mathbb{U}$ implying that

$$x_3 = \mathcal{G}(t, x_1, x_2, p) = \mathcal{G}(t, x_1, x_2, p_0).$$

Thus, there (locally) exists a function $\mathcal{R}$ with

$$\mathcal{R}(t, x_1, x_2) = \mathcal{G}(t, x_1, x_2, p_0).$$

Differentiating (17) in the form

$$\hat{F}_2(t, x_1, x_2, \mathcal{R}(t, x_1, x_2), \mathcal{H}(t, x_1, x_2, p)) \equiv 0$$

with respect to (x_1, x_2), we get (omitting arguments)

$$\begin{aligned} \tfrac{d}{d(x_1,x_2)}\hat{F}_2 &= (Z_{2;x_1,x_2}^T F_\mu + Z_2^T F_{\mu;x_1,x_2}) + (Z_{2;x_3}^T F_\mu + Z_2^T F_{\mu;x_3})\mathcal{R}_{x_1,x_2} \\ &\qquad + (Z_{2;\dot{x},\ldots,x^{(\mu+1)}}^T F_\mu + Z_2^T F_{\mu;\dot{x},\ldots,x^{(\mu+1)}})\mathcal{H}_{x_1,x_2} \\ &= Z_2^T F_{\mu;x_1,x_2} + Z_2^T F_{\mu;x_3}\mathcal{R}_{x_1,x_2} = Z_2^T F_{\mu;x} \begin{bmatrix} I \\ \mathcal{R}_{x_1,x_2} \end{bmatrix} \equiv 0 \end{aligned}$$

on $\mathbb{U}$ such that we can choose T_2 in part 4 of Hypothesis 1 as

$$T_2(t, x_1, x_2) = \begin{bmatrix} I \\ \mathcal{R}_{x_1,x_2}(t, x_1, x_2) \end{bmatrix}. \tag{18}$$

In particular, this means that part 5 of Hypothesis 1 only includes the original variables $(t, x, \dot{x})$. Part 5 also implies that there exists a matrix function Z_1 of size (m, d) with full rank satisfying

$$\operatorname{rank} Z_1^T F_{\dot{x}} T_2 = d \tag{19}$$

on $\mathbb{U}$. Obviously, Z_1 can even be chosen constant.

Summarizing the construction up to now, Hypothesis 1 yields that the original system implies a reduced system (in the original variables) given by

$$\begin{array}{ll} \text{(a)} & \hat{F}_1(t, x_1, x_2, x_3, \dot{x}_1, \dot{x}_2, \dot{x}_3) = 0, \\ \text{(b)} & x_3 - \mathcal{R}(t, x_1, x_2) = 0, \end{array} \tag{20}$$

with $\hat{F}_1 = Z_1^T F$. Eliminating x_3 and $\dot{x}_3$ in (20a) with the help of (20b) and its derivative then leads to

$$\hat{F}_1(t, x_1, x_2, \mathcal{R}(t, x_1, x_2), \dot{x}_1, \dot{x}_2, \mathcal{R}_t(t, x_1, x_2) + \mathcal{R}_{x_1}(t, x_1, x_2)\dot{x}_1 + \mathcal{R}_{x_2}(t, x_1, x_2)\dot{x}_2) = 0.$$

By part 5 of Hypothesis 1 we can assume without loss of generality that this system can (locally) be solved for $\dot{x}_1$ leading to the system

$$\begin{aligned} \dot{x}_1 &= \mathcal{L}(t, x_1, x_2, \dot{x}_2), \\ x_3 &= \mathcal{R}(t, x_1, x_2). \end{aligned} \tag{21}$$

Obviously, in this system $x_2 \in C^1(\mathbb{I}, \mathbb{R}^{n-a-d})$ can be chosen arbitrarily (at least when staying in the domain of definition of $\mathcal{R}$ and $\mathcal{L}$) while the resulting system has locally a unique solution for x_1 and x_3 provided a consistent initial condition is given.

In summary, we have proved the following result.

Theorem 2 *Let F in (1) be sufficiently smooth and satisfy Hypothesis 1 with μ, a, d, v. Then every solution of (1) also solves the reduced problems (20) and (21) consisting of d differential and a algebraic equations.*

So far, we have not used the quantity v. This quantity measures the number of equations in the original system that gives rise to trivial equations $0 = 0$, i. e., it counts the number of redundancies in the system. Together with a and d it gives a complete classification of the m equations into d differential equations, a algebraic equations and v trivial equations. Of course, trivial equations can be simply removed without altering the solution set. Omitting part 3 of Hypothesis 1 would mean that a given problem may satisfy the modified hypothesis for different values of a and d.

Example 2 Setting $x_3 = 0$ in (5) of Example 1 gives the problem

$$F(t, x, \dot{x}) = \begin{bmatrix} \dot{x}_2 \\ \log x_2 \end{bmatrix} = 0, \tag{22}$$

with $m = 2$ and $n = 2$. Note that here x_1, x_2 denote the components of x and not the splitting of x used in the above theoretical construction. To check Hypothesis 1 for $\mu = 0$ we consider the set

$$\mathbb{L}_0 = \{(t, x_1, x_2, \dot{x}_1, \dot{x}_2) \mid x_2 = 1,\ \dot{x}_2 = 0\}.$$

Obviously, $\mathbb{L}_0$ is a manifold parametrized by $(t, x_1, \dot{x}_1)$. Furthermore, we have

$$F_{0;\dot{x}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad F_{0;x} = \begin{bmatrix} 0 & 0 \\ 0 & x_2^{-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

on $\mathbb{L}_0$. Thus,

$$\operatorname{rank} F_{0;x,\dot{x}} = 2, \quad \operatorname{corank} F_{0;x,\dot{x}} = 0, \quad \operatorname{rank} F_{0;\dot{x}} = 1.$$

With $Z_2^T = [\,0\ \ 1\,]$, we then obtain

$$\operatorname{rank} Z_2^T F_{0;x} = \operatorname{rank}[\,0\ \ 1\,] = 1,$$

and with $T_2^T = [\,1\ \ 0\,]$ finally

$$\operatorname{rank} F_{\dot{x}} T_2 = 0.$$

Hence, we get the quantities $r = 2$, $v = 0$, $a = 1$, and $d = 0$. Hypothesis 1 is not satisfied, since $d \neq m - a - v = 1$. If we would drop part 3 of Hypothesis 1, there would be no condition on v and we could simply choose $v = 1$ to satisfy all remaining requirements. To check Hypothesis 1 for $\mu = 1$ we must deal with $F_1 = 0$, which consists of the equations

$$\dot{x}_2 = 0, \quad \log x_2 = 0, \quad \ddot{x}_2 = 0, \quad \frac{\dot{x}_2}{x_2} = 0.$$

The set

$$\mathbb{L}_1 = \{(t, x_1, x_2, \dot{x}_1, \dot{x}_2, \ddot{x}_1, \ddot{x}_2) \mid x_2 = 1,\ \dot{x}_2 = 0,\ \ddot{x}_2 = 0\}$$

is a manifold parametrized by $(t, x_1, \dot{x}_1, \ddot{x}_1)$. Furthermore, we have

$$F_{1;\dot{x},\ddot{x}} = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & x_2^{-1} & 0 & 0 \end{array}\right] = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{array}\right]$$

and

$$F_{1;x} = \left[\begin{array}{cc} 0 & 0 \\ 0 & x_2^{-1} \\ \hline 0 & 0 \\ 0 & -x_2^{-2}\dot{x}_2 \end{array}\right] = \left[\begin{array}{cc} 0 & 0 \\ 0 & 1 \\ \hline 0 & 0 \\ 0 & 0 \end{array}\right]$$

on $\mathbb{L}_1$. Thus,

$$\operatorname{rank} F_{1;x,\dot{x},\ddot{x}} = 3, \quad \operatorname{corank} F_{1;x,\dot{x},\ddot{x}} = 1, \quad \operatorname{rank} F_{1;\dot{x},\ddot{x}} = 2.$$

Proceeding as above, we compute

$$Z_2^T = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 \end{array}\right], \quad T_2 = \left[\begin{array}{c} 1 \\ 0 \end{array}\right],$$

and

$$\operatorname{rank} Z_2^T F_{0;x} = \operatorname{rank} \left[\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array}\right] = 1, \quad \operatorname{rank} F_{\dot{x}} T_2 = \operatorname{rank} \left[\begin{array}{c} 0 \\ 0 \end{array}\right] = 0.$$

Hence, Hypothesis 1 is satisfied with $\mu = 1$, $r = 2$, $v = 1$, $a = 1$, and $d = 0$.

Theorem 2 states that every (sufficiently smooth) solution x of the original system (1) also solves the reduced systems (20) and (21). To show that the reduced systems reflect (at least locally) the properties of the original system concerning solvability and structure of the solution set, we need the converse direction of this statement. The following theorem gives sufficient conditions.

Theorem 3 *Let F in (1) be sufficiently smooth and satisfy Hypothesis 1 with μ, a, d, v and with $\mu+1$ (replacing μ), a, d, v. Let $z^0_{\mu+1} \in \mathbb{L}_{\mu+1}$ be given and let the parametrization p in (15) for $F_{\mu+1}$ include $\dot{x}_2$. Then, for every function $x_2 \in C^1(\mathbb{I}, \mathbb{R}^{n-a-d})$ with $x_2(t_0) = x_{20}$, $\dot{x}_2(t_0) = \dot{x}_{20}$ the reduced problem (21) has unique solutions x_1 and x_3 satisfying $x_1(t_0) = x_{10}$. Moreover, these together locally solve the original problem.*

Proof. By assumption, there exists (locally with respect to $z^0_{\mu+1} \in \mathbb{L}_{\mu+1}$) a parametrization (t, x_1, x_2, p), where p is chosen out of $(\dot{x}, \ldots, x^{(\mu+2)})$, with

$$F_{\mu+1}(t, x_1, x_2, \mathcal{R}(t, x_1, x_2), \mathcal{H}(t, x_1, x_2, p)) \equiv 0.$$

This includes the equation

$$F_\mu(t,x_1,x_2,\mathcal{R}(t,x_1,x_2),\mathcal{H}(t,x_1,x_2,p)) \equiv 0 \tag{23}$$

with trivial dependence on $x^{(\mu+2)}$ as well as

$$\frac{d}{dt}F_\mu(t,x_1,x_2,\mathcal{R}(t,x_1,x_2),\mathcal{H}(t,x_1,x_2,p)) \equiv 0. \tag{24}$$

Equation (23) implies that (omitting arguments)

$$\begin{array}{l} F_{\mu;t} + F_{\mu;x_3}\mathcal{R}_t + F_{\mu;\dot{x},\ldots,x^{(\mu+2)}}\mathcal{H}_t \equiv 0, \\ F_{\mu;x_1,x_2} + F_{\mu;x_3}\mathcal{R}_{x_1,x_2} + F_{\mu;\dot{x},\ldots,x^{(\mu+2)}}\mathcal{H}_{x_1,x_2} \equiv 0, \\ F_{\mu;\dot{x},\ldots,x^{(\mu+2)}}\mathcal{H}_p \equiv 0. \end{array} \tag{25}$$

The relation $\frac{d}{dt}F_\mu = 0$ has the form

$$F_{\mu;t} + F_{\mu;x_1}\dot{x}_1 + F_{\mu;x_2}\dot{x}_2 + F_{\mu;x_3}\dot{x}_3 + F_{\mu;\dot{x},\ldots,x^{(\mu+1)}}\left[\begin{array}{c} \ddot{x} \\ \vdots \\ x^{(\mu+2)} \end{array}\right] = 0.$$

Inserting the parametrization yields that (24) can be written as

$$F_{\mu;t} + F_{\mu;x_1}\mathcal{H}_1 + F_{\mu;x_2}\mathcal{H}_2 + F_{\mu;x_3}\mathcal{H}_3 + F_{\mu;\dot{x},\ldots,x^{(\mu+1)}}\mathcal{H}_4 \equiv 0,$$

where $\mathcal{H}_i$, $i = 1,\ldots,4$, are the parts of $\mathcal{H}$ corresponding to $\dot{x}_1$, $\dot{x}_2$, $\dot{x}_3$, and the remaining variables, respectively. Multiplication with Z_2^T (corresponding to Hypothesis 1 with μ, a, d, v) gives

$$Z_2^T F_{\mu;t} + Z_2^T F_{\mu;x_1}\mathcal{H}_1 + Z_2^T F_{\mu;x_2}\mathcal{H}_2 + Z_2^T F_{\mu;x_3}\mathcal{H}_3 \equiv 0.$$

Inserting the relations (25) and observing that $Z_2^T F_{\mu;x_3}$ is nonsingular, we find

$$Z_2^T F_{\mu;x_3}(\mathcal{H}_3 - \mathcal{R}_t - \mathcal{R}_{x_1}\mathcal{H}_1 - \mathcal{R}_{x_2}\mathcal{H}_2) \equiv 0,$$

or

$$\mathcal{H}_3 = \mathcal{R}_t + \mathcal{R}_{x_1}\mathcal{H}_1 + \mathcal{R}_{x_2}\mathcal{H}_2,$$

that is

$$\dot{x}_3 = \mathcal{R}_t + \mathcal{R}_{x_1}\dot{x}_1 + \mathcal{R}_{x_2}\dot{x}_2.$$

In summary, the derivative array $F_{\mu+1} = 0$ implies that

$$\begin{array}{ll} \text{(a)} & Z_1^T F(t,x_1,x_2,x_3,\dot{x}_1,\dot{x}_2,\dot{x}_3) = 0, \\ \text{(b)} & x_3 = \mathcal{R}(t,x_1,x_2), \\ \text{(c)} & \dot{x}_3 = \mathcal{R}_t(t,x_1,x_2) + \mathcal{R}_{x_1}\dot{x}_1(t,x_1,x_2) + \mathcal{R}_{x_2}(t,x_1,x_2)\dot{x}_2. \end{array} \tag{26}$$

Elimination of x_3 and $\dot{x}_3$ from (26a) gives

$$\dot{x}_1 = \mathcal{L}(t,x_1,x_2,\dot{x}_2).$$

In particular, $\dot{x}_1$ and $\dot{x}_3$ are not part of the parametrization.

Since $\dot{x}_2$ is part of p, the following construction is possible. Let $x_2 = x_2(t)$ and $\dot{x}_2 = \dot{x}_2(t)$. Let $p = p(t)$ be arbitrary but consistent to the choice of $\dot{x}_2$ and to the initial value $z^0_{\mu+1}$. Finally, let $x_1 = x_1(t)$ and $x_3 = x_3(t)$ be the solution of the initial value problem

$$\begin{array}{l} Z_1^T F(t, x_1, x_2(t), x_3, \dot{x}_1, \dot{x}_2(t), \dot{x}_3) = 0, \quad x_1(t_0) = x_{10} \\ x_3 = \mathcal{R}(t, x_1, x_2(t)). \end{array}$$

Although $\dot{x}_1$ and $\dot{x}_3$ are not part of the parametrization, we automatically get $\dot{x}_1 = \dot{x}_1(t)$ and $\dot{x}_3 = \dot{x}_3(t)$. Thus, we have

$$F_{\mu+1}(t, x_1(t), x_2(t), x_3(t), \dot{x}_1(t), \dot{x}_2(t), \dot{x}_3(t), \mathcal{H}_4(t, x_1(t), x_2(t), p(t)) \equiv 0,$$

for all t in a neighborhood of t_0, or

$$F(t, x_1(t), x_2(t), x_3(t), \dot{x}_1(t), \dot{x}_2(t), \dot{x}_3(t)) \equiv 0$$

for the first block. □

Corollary 4 *Let F in (1) be sufficiently smooth and satisfy Hypothesis 1 with μ, a, d, v and with $\mu+1$ (replacing μ), a, d, v and assume that $a + d = n$. For every $z^0_{\mu+1} \in \mathbb{L}_{\mu+1}$ the reduced problem (21) has a unique solution satisfying the initial value given by $z^0_{\mu+1}$. Moreover, this solution locally solves the original problem.*

Proof. Since $a + d = n$, there is no part x_2 of x in the above construction. □

The above corollary especially applies to the case of regular problems as treated in [26], where we have $m = n$ and $v = 0$. Together with the observation that every (sufficiently smooth) solution also solves the reduced problem, we have now found sufficient conditions that guarantee that original problem and reduced problem (locally) show the same behaviour concerning solvability and the structure of the solution set.

Remark 1 Let the assumptions of Theorem 3 hold and let x_{20} and $\dot{x}_{20}$ be the part of $z^0_{\mu+1} \in \mathbb{L}_{\mu+1}$ belonging to x_2 and $\dot{x}_2$. If $\tilde{x}_{20}$ and $\dot{\tilde{x}}_{20}$ are sufficiently close to x_{20} and $\dot{x}_{20}$, they are part of a $\tilde{z}^0_{\mu+1} \in \mathbb{L}_{\mu+1}$ close to $z^0_{\mu+1}$ and we can apply Theorem 3 with $z^0_{\mu+1}$ replaced by $\tilde{z}^0_{\mu+1}$.

Remark 2 Note that in Theorem 3 we can drop the assumption that $\dot{x}_2$ is part of the parameters if we know from the structure of the problem that $\mathcal{L}$ in (21) does not depend on $\dot{x}_2$. In particular, this is the case if we can choose the splitting (x_1, x_2, x_3) in such a way that the original problem does not depend on $\dot{x}_2$ and on components of $\dot{x}_3$ that depend on $\dot{x}_2$. An important consequence of this special case is that we need not to require the initial condition $\dot{x}_2(t_0) = \dot{x}_{20}$. This also applies to Remark 1.

Remark 3 Although we must deal with $F_{\mu+1}$ in order to show that the solutions of the reduced problem also solve the original problem, it is sufficient to consider F_μ only in order to obtain the reduced problem and to solve it. This could already be observed in the linear case, see [25]. Compare also with the numerical procedures in Section 5.

Remark 4 The reduced problems (20) and (21) may already follow from $F_\ell = 0$ with $\ell < \mu$, although μ is chosen as small as possible. This occurs in cases where further differentiations only lead to trivial equations $0 = 0$ (when consistency is guaranteed). To check the consistency of the model, however, it is still necessary to consider $F_\mu = 0$.

Example 3 Consider the problem of Example 2. The reduced problem simply consists of $\log x_2 = 0$ and is already implied by $F_0 = 0$. The same holds for the slightly modified problem

$$\dot{x}_2 = 1, \quad \log x_2 = 0.$$

Observe that the corresponding set $\mathbb{L}_0$ is nonempty. Differentiating once gives

$$\ddot{x}_2 = 0, \quad x_2^{-1}\dot{x}_2 = 0$$

implying the contradiction $\dot{x}_2 = 0$. Thus, $\mathbb{L}_1$ is empty and the modified problem is not solvable.

4 Application to control problems

In this section we apply the results from the previous section to control problems of the form (2). In the linear case this has been the topic of numerous publications [3, 4, 5, 6, 32, 33, 27]. In particular in [5, 6, 32, 33, 27] the general case of nonsquare control problems has been discussed concerning solvability, regularizability, model consistency and conditions have been derived that guarantee that the system can be regularized by state or output feedback or how it can be reinterpreted as a square strangeness-free system. To do this, redundancies are removed, free variables are reinterpreted as controls and fixed controls are reinterpreted as state variables. In the nonlinear case we have already shown under which circumstances redundancies can be removed, but we will assume in the following that a reinterpretation of variables is not necessary, i.e., controls are variables that can be freely chosen and state variables are variables that are determined from the system, once a control has been chosen. In the behaviour approach such an assumption is not really necessary but it simplifies the notation which is already quite involved.

Consider the control problem without the output equation, i. e., $F(t, \xi, u, \dot{\xi}) = 0$ with $\xi \in \mathbb{R}^{n_\xi}$, $u \in \mathbb{R}^{n_u}$ and $n = n_\xi + n_u$. In a behaviour framework, we set

$$x = \begin{bmatrix} \xi \\ u \end{bmatrix}$$

and apply the theory of the previous section. This gives locally a reduced problem of the form

$$\begin{aligned} \hat{F}_1(t,\xi,u,\dot{\xi}) &= 0, \\ \hat{F}_2(t,\xi,u) &= 0 \end{aligned} \tag{27}$$

corresponding to (20). To perform the next steps of the construction would require to split x into (x_1, x_2, x_3) where each part may consist of components of both ξ and u. To avoid such a splitting we proceed as follows. Starting from (20) in the form

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

Hypothesis 1 yields (without arguments)

$$\hat{F}_{2;x}T_2 = 0, \quad \operatorname{rank} T_2 = n-a, \quad \operatorname{rank} \hat{F}_{1;\dot{x}} = d.$$

Choosing T_2' such that $[\, T_2' \;\, T_2 \,]$ is nonsingular, we find

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

Thus, the given matrix has full row rank. In the present context, this means that the $(d+a,n)$-matrix

$$\begin{bmatrix} \hat{F}_{1;\dot{\xi}} & 0 \\ \hat{F}_{2;\xi} & \hat{F}_{2;u} \end{bmatrix} \tag{28}$$

has full row rank. Observe that in general fixing a control u does not give a regular strangeness-free reduced problem (in the sense of [26]), since

$$\begin{bmatrix} \hat{F}_{1;\dot{\xi}} \\ \hat{F}_{2;\xi} \end{bmatrix}$$

may be singular. An immediate question is whether it is possible to choose a control such that the resulting reduced problem is regular and strangeness-free. Necessarily, we must have $d+a = n_\xi$. As in the linear case (see [27]) we consider state feedbacks and output feedbacks. In the nonlinear case a state feedback may have the form

$$u = K(t,\xi) \tag{29}$$

leading to a closed loop reduced problem

$$\begin{aligned} \hat{F}_1(t,\xi,K(t,\xi),\dot{\xi}) &= 0, \\ \hat{F}_2(t,\xi,K(t,\xi)) &= 0. \end{aligned} \tag{30}$$

The condition for this system to be regular and strangeness-free reads

$$\begin{bmatrix} \hat{F}_{1;\dot{\xi}} \\ \hat{F}_{2;\xi} + \hat{F}_{2;u}K_\xi \end{bmatrix} \quad \text{nonsingular.}$$

Since (28) has full rank, the existence of a suitable $\tilde{K} = K_\xi$ follows from the theory for linear problems with constant coefficients. Thus a possible state feedback is given by

$$u(t) = \tilde{K}\xi(t) + w(t), \tag{31}$$

where the function w can be used to satisfy initial conditions of the form

$$u^{(\ell)}(t_0) = \tilde{K}\xi_0^{(\ell)} + w^{(\ell)}(t_0) = u_0^{(\ell)}. \tag{32}$$

Hence, we have proved the following theorem.

Theorem 5 *Suppose that the control problem (2) in behaviour form satisfies Hypothesis 1 with μ, a, d, v and assume that $d + a = n_\xi$. Then there (locally) exists a state feedback $u = K(t, \xi)$ satisfying $u_0 = K(t_0, \xi_0)$ and $\dot{u}_0 = K_t(t_0, \xi_0) + K_\xi(t_0, \xi_0)\dot{\xi}_0$ such that the closed loop reduced problem is regular and strangeness-free.*

Corollary 6 *Suppose that the control problem (2) in behaviour form satisfies Hypothesis 1 with μ, a, d, v and with $\mu + 1$ (replacing μ), a, d, v and assume that $d + a = n_\xi$. Furthermore, let u be a control in the sense that u and $\dot{u}$ can be chosen as part of the parametrization of $\mathbb{L}_{\mu+1}$ at $z^0_{\mu+1} \in \mathbb{L}_{\mu+1}$. Let $u = K(t, \xi)$ be a state feedback which satisfies the initial conditions $u_0 = K(t_0, \xi_0)$ and $\dot{u}_0 = K_t(t_0, \xi_0) + K_\xi(t_0, \xi_0)\dot{\xi}_0$ and yields a regular and strangeness-free closed loop reduced system. Then, the closed loop reduced problem has a unique solution satisfying the initial values given by $z^0_{\mu+1}$. Moreover, this solution locally solves the closed loop problem*

$$F(t, \xi, K(t, \xi), \dot{\xi}) = 0.$$

Proof. The proof follows the lines of that of Theorem 3. $\square$

Example 4 Consider the control problem (4) of Example 1 and the corresponding behaviour system (5). To check Hypothesis 1 for $\mu = 0$ we use

$$x = \begin{bmatrix} \xi_1 \\ \xi_2 \\ u \end{bmatrix}.$$

The set

$$\mathbb{L}_0 = \{(t, \xi_1, \xi_2, u, \dot{\xi}_1, \dot{\xi}_2, \dot{u}) \mid \xi_2 = \exp(-\sin u),\ \dot{\xi}_2 = 0\}.$$

is a manifold parametrized by $(t, \xi_1, u, \dot{\xi}_1, \dot{u})$. Furthermore, we have

$$F_{0;\dot{x}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad F_{0;x} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \xi_2^{-1} & \cos u \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \exp(\sin u) & \cos u \end{bmatrix}$$

on $\mathbb{L}_0$. Thus,

$$\operatorname{rank} F_{0;x,\dot{x}} = 2, \quad \operatorname{corank} F_{0;x,\dot{x}} = 0, \quad \operatorname{rank} F_{0;\dot{x}} = 1.$$

With $Z_2^T = [\,0\ \ 1\,]$, we then obtain

$$\operatorname{rank} Z_2^T F_{0;x} = \operatorname{rank}[\,0\ \ \exp(\sin u)\ \ \cos u\,] = 1, \quad T_2 = \begin{bmatrix} 1 & 0 \\ 0 & -\cos u \\ 0 & \exp(\sin u) \end{bmatrix},$$

and finally

$$\operatorname{rank} F_{\dot{x}} T_2 = \operatorname{rank} \begin{bmatrix} 0 & -\cos u \\ 0 & 0 \end{bmatrix} = 1,$$

when we restrict u to a neighborhood of zero. Hence, Hypothesis 1 is satisfied with $\mu = 0$, $r = 2$, $v = 0$, $a = 1$, and $d = 1$. For $z_0^0 = (0, 0, 1, 0, 0, 1, 0)$ we can choose $Z_1^T = [\,1\ \ 0\,]$ to obtain the reduced problem

$$\dot{\xi}_2 = 0, \quad \log \xi_2 + \sin u = 0.$$

Note that the reduced problem here coincides with the original problem due to its special form (we have $\mu = 0$ and do not need to apply any transformations to separate the algebraic equations) and due to the special choice for Z_1^T. Fixing the control u according to $u = 0$ gives a closed loop system that is not regular and strangeness-free. Indeed, it satisfies Hypothesis 1 only for $\mu = 1$ and it even includes a trivial equation due to a redundancy, cp. Example 2. To get a regular and strangeness-free closed-loop reduced problem, we look for a regularizing state feedback. Since

$$\begin{bmatrix} \hat{F}_{1;\dot{\xi}} & 0 \\ \hat{F}_{2;\xi} & \hat{F}_{2;u} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & \xi_2^{-1} & \cos u \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

at z_0^0, we can choose $\tilde{K} = [\,1\ \ 0\,]$ or $u = \xi_1$ observing the initial values given by z_0^0. The corresponding closed loop reduced problem is given by

$$\dot{\xi}_2 = 0, \quad \log \xi_2 + \sin \xi_1 = 0.$$

By construction, it is regular and strangeness-free near the initial value given by z_0^0. For $\xi_1(0) = 0$, we particularly get the unique solution $\xi_1(t) = 0$, $\xi_2(t) = 1$.

We turn now to control problems that include the output equation (3), i. e., $F(t, \xi, u, \dot{\xi}) = 0$ together with $y = G(t, \xi)$, where $\xi \in \mathbb{R}^{n_\xi}$, $u \in \mathbb{R}^{n_u}$, $y \in \mathbb{R}^{n_y}$ and $n = n_\xi + n_u + n_y$. In a behaviour framework, we set

$$x = \begin{bmatrix} \xi \\ u \\ y \end{bmatrix}$$

and again apply the theory of the previous section. Due to the explicit form of the output equation, it is obvious that it becomes part of the algebraic constraints and does not affect the other constraints, cp. the linear case in [27]. Therefore, the reduced problem has the form

$$\begin{array}{l} \hat{F}_1(t,\xi,u,\dot{\xi}) = 0, \\ \hat{F}_2(t,\xi,u) = 0, \\ y = G(t,\xi). \end{array} \tag{33}$$

If we consider output feedbacks

$$u = K(t,y), \tag{34}$$

the closed loop reduced problem has the form

$$\begin{array}{l} \hat{F}_1(t,\xi,K(t,G(t,\xi)),\dot{\xi}) = 0, \\ \hat{F}_2(t,\xi,K(t,g(t,\xi))) = 0. \end{array} \tag{35}$$

The condition for this system to be regular and strangeness-free reads

$$\left[\begin{array}{c} \hat{F}_{1;\dot{\xi}} \\ \hat{F}_{2;\xi} + \hat{F}_{2;u}K_yG_\xi \end{array}\right] = \left[\begin{array}{cc} \hat{F}_{1;\dot{\xi}} & 0 \\ \hat{F}_{2;\xi} & \hat{F}_{2;u} \end{array}\right] \left[\begin{array}{c} I \\ K_yG_\xi \end{array}\right] \quad \text{nonsingular.} \tag{36}$$

Note that we get back the state feedback case if $y = \xi$. To guarantee that condition (36) holds for some choice of K_y, we need an extra condition which we can check locally via the following procedure, see Algorithm 1 in [27]. As there, this algorithm directly allows the construction of a suitable linear output feedback that satisfies the above regularity condition.

Algorithm 1 Let the Jacobians $E_1 = \hat{F}_{1;\dot{\xi}}$, $A_2 = \hat{F}_{2;\xi}$, $B_2 = \hat{F}_{2;u}$ and $C = G_\xi$ of the reduced system corresponding to $z_\mu^0 \in \mathbb{L}_\mu$ be given.

1. Determine an orthogonal matrix $Q = [\, Q_1 \;\; Q_2 \,]$ such that

$$E_1[\, Q_1 \;\; Q_2 \,] = [\, E_{11} \;\; 0 \,],$$

   where E_{11} has size (d,d) and is nonsingular.

2. Determine orthogonal matrices $U = [\, U_1 \;\; U_2 \,]$ and $V = [\, V_1 \;\; V_2 \,]$ such that

$$U^T A_2 Q_2 V = \left[\begin{array}{cc} A_{22} & 0 \\ 0 & 0 \end{array}\right],$$

   where A_{22} is of size $(\hat{a},\hat{a})$ and nonsingular. Set $\phi = a - \hat{a}$ and check if $\operatorname{rank} U_2^T B_2 = \phi$.

3. Determine the rank ω of CQ_2V_2. In particular, determine an orthogonal matrix $W = [\, W_1 \;\; W_2 \,]$ such that

$$CQ_2V_2W = [\, C_3 \;\; 0 \,],$$

   where C_3 has full column rank ω.

Theorem 7 *Suppose that the output control problem, consisting of (2) and (3) in behaviour form, satisfies Hypothesis 1 with μ, a, d, v and assume that $d+a = n_\xi$ as well as $\phi = \omega$ for the quantities determined by Algorithm 1. Then there (locally) exists an output feedback $u = K(t, y)$ satisfying $u_0 = K(t_0, y_0)$ and $\dot{u}_0 = K_t(t_0, y_0) + K_\xi(t_0, y_0)\dot{y}_0$ such that the closed loop reduced problem is regular and strangeness-free.*

Proof. Under the given assumptions, the linear theory (involving Algorithm 1) yields a suitable matrix $\tilde{K} = K_y$ such that (36) holds. The claim then follows for the linear output feedback

$$u(t) = \tilde{K}y(t) + w(t),$$

where the function w is used to satisfy the given initial conditions. ☐

Corollary 8 *Suppose that the output control problem, consisting of (2) and (3) in behaviour form, satisfies Hypothesis 1 with μ, a, d, v and with $\mu + 1$ (replacing μ), a, d, v and assume that $d+a = n_\xi$ as well as $\phi = \omega$ for the quantities determined by Algorithm 1. Furthermore, let u be a control in the sense that u and $\dot{u}$ can be chosen as part of the parametrization of $\mathbb{L}_{\mu+1}$ at $z^0_{\mu+1} \in \mathbb{L}_{\mu+1}$. Let $u = K(t, y)$ be an output feedback which satisfies the initial conditions $u_0 = K(t_0, y_0)$ and $\dot{u}_0 = K_t(t_0, y_0) + K_\xi(t_0, y_0)\dot{y}_0$ and yields a regular and strangeness-free closed loop reduced system. Then, the closed loop reduced problem has a unique solution satisfying the initial values given by $z^0_{\mu+1}$. Moreover, this solution locally solves the closed loop problem*

$$F(t, \xi, K(t, G(t, \xi)), \dot{\xi}) = 0.$$

Proof. Again, the proof follows the lines of that of Theorem 3. ☐

Remark 5 As in the case of the previous section, it is sufficient to consider F_μ in order to compute the desired regularizing state or output feedback and the solution of the closed loop system.

Remark 6 Although all obtained results were of local nature, they can be globalized as it can be done in the case of ordinary differential equations (see, e. g., [21, Th. I.7.4]). Like there, we can continue the process (under the assumption of sufficient smoothness) until we reach the boundary of $\mathbb{L}_\mu$ or $\mathbb{L}_{\mu+1}$, respectively. Note that this may happen in finite time.

Remark 7 Suppose that for a given control problem (2) the variable ξ can be split into (ξ_1, ξ_2) in such a way that the reduced problem (27) can be transformed to

$$\begin{aligned} \dot{\xi}_1 &= \mathcal{L}(t, \xi_1, u), \\ \xi_2 &= \mathcal{R}(t, \xi_1, u) \end{aligned}$$

according to (21). Then for every u with $u(t_0)$ sufficiently close to u_0 the closed loop reduced problem obviously is regular and strangeness-free. Due to the structure of the problem (cp. Remark 2), we do not need to require that $\dot{u}$ is part of the parameters in order to get the results of Corollaries 6 and 8. Accordingly, we do not need to require that $\dot{u}(t_0) = \dot{u}_0$.

Example 5 A control problem for a multibody system has the form (see, e. g., [34])

$$\begin{aligned}
&\dot{p} = q, \\
&M(p)\dot{q} = f(t,p,q,u) + g_p(p)^T\lambda, \quad p \in \mathbb{R}^{n_p},\ \lambda \in \mathbb{R}^{n_\lambda} \\
&g(p) = 0,
\end{aligned}$$

since the control typically acts via external forces. Assuming that $g_p(p)$ has full row rank and $M(p)$ is symmetric and positive definite, Hypothesis 1 is satisfied with $\mu = 2$, $d = 2(n_p - n_\lambda)$, $a = 3n_\lambda$, and $v = 0$, provided the model is consistent according to $\mathbb{L}_\mu \neq \emptyset$. The corresponding reduced problem has the form

$$\begin{aligned}
&Z_{11}^T(\dot{p} - q) = 0, \\
&Z_{12}^T(M(p)\dot{q} - f(t,p,q,u) - g_p(p)^T\lambda) = 0, \\
&g(p) = 0, \\
&g_p(p)q = 0, \\
&g_{pp}(q,q) + g_p(p)M(p)^{-1}[f(t,p,q,u) + g_p(p)^T\lambda] = 0
\end{aligned}$$

and can be shown to be regular and strangeness-free for given u near the initial value. Due to the assumptions, we can split $p = (p_1, p_2)$ and $q = (q_1, q_2)$ such that in the notation of the previous section

$$x_1 = (p_1, q_1), \quad x_2 = u, \quad x_3 = (p_2, q_2, \lambda)$$

is a possible choice. The special structure of the reduced problem implies that from $\dot{x}_3$ only $\dot{\lambda}$ may depend on $\dot{u}$. Thus, Remarks 2 and 7 apply.

5 Numerical methods

The theoretical results of the previous two sections directly imply numerical methods for the computation of the desired solutions. In the general case of Section 3 we can use the following numerical procedures.

To compute a consistent initial value at time t_0, i. e., a value x_0 that satisfies the algebraic constraints, we must solve

$$F_\mu(t_0, x_0, \dot{x}_0, \ldots, x_0^{(\mu+1)}) = 0 \tag{37}$$

for $(x_0, \dot{x}_0, \ldots, x_0^{(\mu+1)})$. The classical approach to solve such systems is the Gauß-Newton method. For a nonlinear problem $\mathcal{F}(c) = 0$ it generates a sequence c_k of approximations starting with an initial guess c_0 by

$$c_{k+1} = c_k - \mathcal{F}_c(c_k)^{-}\mathcal{F}(c_k), \tag{38}$$

where $\mathcal{F}_c(c_k)^-$ denotes a convenient (outer or left) generalized inverse (see, e. g., [7]) of $\mathcal{F}_c(c_k)$. Due to the required consistency of the equations, i. e., $\mathbb{L}_\mu \neq \emptyset$, we expect (for a sufficiently good initial guess) superlinear convergence of the Gauß-Newton method to a solution of the system. For more details see Remark 8 below.

To perform an integration step from t_0 to $t_1 = t_0 + h$ we first determine a projection P that selects a suitable set of components from x which can serve as controls (in the notation of Section 3 this was x_2) and a possible Z_1 at z_μ^0 according to Hypothesis 1. For a suitable control u satisfying

$$u(t_0) = Px_0, \quad \dot{u}(t_0) = P\dot{x}_0,$$

we combine the equation $F_\mu(z_\mu) = 0$, which implies that the algebraic constraints are fulfilled, with the discretized differential equations. Denoting by $D_h x$ a BDF-discretization of $\dot{x}$ (see, e. g., [1]), we obtain

$$\begin{array}{ll} F_\mu(t_1, x_1, \dot{x}_1, \ldots, x_1^{(\mu+1)}) = 0, & Px_1 = u(t_1), \\ Z_1^T F(t_1, x_1, P\dot{x}_1 + (I-P)D_h x_1) = 0, & P\dot{x}_1 = \dot{u}(t_1), \end{array} \tag{39}$$

which must be solved for $(x_1, \dot{x}_1, \ldots, x_1^{(\mu+1)})$. Again we may apply the Gauß-Newton method and expect superlinear convergence. Note that the quality of an initial guess is here not crucial, since we can simple reduce the stepsize h.

In the case of a control problem without output equation, we solve

$$F_\mu(t_0, \xi_0, u_0, \dot{\xi}_0, \dot{u}_0, \ldots, \xi_0^{(\mu+1)}, u_0^{(\mu+1)}) = 0 \tag{40}$$

for $(\xi_0, u_0, \dot{\xi}_0, \dot{u}_0, \ldots, \xi_0^{(\mu+1)}, u_0^{(\mu+1)})$ to obtain consistent initial values. Then, we determine Z_1 as above and a suitable $\tilde{K}$ yielding a regularizing state feedback as described in the previous section and set $w = u_0 - \tilde{K}\xi_0$. Finally, we perform an integration step by solving

$$\begin{array}{l} F_\mu(t_1, \xi_1, \tilde{K}\xi_1 + w, \dot{\xi}_1, \dot{u}_1, \ldots, \xi_1^{(\mu+1)}, u_1^{(\mu+1)}) = 0, \\ Z_1^T F(t_1, \xi_1, \tilde{K}\xi_1 + w, D_h\xi_1) = 0 \end{array} \tag{41}$$

for $(\xi_1, \dot{\xi}_1, \dot{u}_1, \ldots, \xi_1^{(\mu+1)}, u_1^{(\mu+1)})$. Under the assumptions of Theorem 5, the Gauß-Newton method will show superlinear convergence for sufficiently small h.

Including the output equation, we accordingly solve

$$F_\mu(t_0, \xi_0, u_0, y_0, \dot{\xi}_0, \dot{u}_0, \dot{y}, \ldots, \xi_0^{(\mu+1)}, u_0^{(\mu+1)}, y_0^{(\mu+1)}) = 0 \tag{42}$$

for $(\xi_0, u_0, y_0, \dot{\xi}_0, \dot{u}_0, \dot{y}_0, \ldots, \xi_0^{(\mu+1)}, u_0^{(\mu+1)}, y_0^{(\mu+1)})$ to obtain consistent initial values. We again determine Z_1 and a suitable $\tilde{K}$ yielding a regularizing output feedback and set $w = u_0 - \tilde{K}y_0$. Here we must solve

$$\begin{array}{l} F_\mu(t_1, \xi_1, \tilde{K}y_1 + w, y_1, \dot{\xi}_1, \dot{u}_1, \dot{y}_1, \ldots, \xi_1^{(\mu+1)}, u_1^{(\mu+1)}, y_1^{(\mu+1)}) = 0, \\ Z_1^T F(t_1, \xi_1, \tilde{K}y_1 + w, D_h\xi_1) = 0 \end{array} \tag{43}$$

for $(\xi_1, y_1, \dot{\xi}_1, \dot{u}_1, \dot{y}_1, \ldots, \xi_1^{(\mu+1)}, u_1^{(\mu+1)}, y_1^{(\mu+1)})$. Due to the explicit form of the output equation, we can remove it and all its derivatives from F_μ. Denoting the resulting function by $\tilde{F}_\mu$, it is sufficient to solve

$$\begin{array}{l} \tilde{F}_\mu(t_1, \xi_1, \tilde{K}G(t_1, \xi_1) + w, \dot{\xi}_1, \dot{u}_1, \ldots, \xi_1^{(\mu+1)}, u_1^{(\mu+1)}) = 0, \\ Z_1^T F(t_1, \xi_1, \tilde{K}G(t_1, \xi_1) + w, D_h \xi_1) = 0 \end{array} \tag{44}$$

for $(\xi_1, \dot{\xi}_1, \dot{u}_1, \ldots, \xi_1^{(\mu+1)}, u_1^{(\mu+1)})$ and we may determine $(y_1, \dot{y}_1, \ldots, y_1^{(\mu+1)})$ by the explicit formulas given by the output equation and its derivatives. Still, we expect superlinear convergence due to Theorem 7 for sufficiently small h.

Having performed an integration step, we always end up with a new consistent value on $\mathbb{L}_\mu$, since in all cases the equation $F_\mu(z_\mu) = 0$ is part of the numerical procedure. Thus, we can iteratively proceed with the integration giving at least piecewise smooth regularizing controls and associated solutions.

Remark 8 In order to perform the Gauß-Newton iteration (38) we must specify how we choose the generalized inverse $\mathcal{F}_c(c)^-$. Since we know the rank of the Jacobian at the desired solution (say r as for (37)), we can proceed as follows. We compute a QR-decomposition with column pivoting of $\mathcal{F}_c(c_0)$ of the form

$$Q_0^T \mathcal{F}_c(c_0)\Pi = \left[\begin{array}{cc} R_0 & S_0 \\ 0 & \Delta_0 \end{array} \right],$$

where Q_0 is orthogonal, R_0 is nonsingular with rank r, and Π is a permutation matrix. For c sufficiently close to c_0, we can determine a QR-decomposition of $\mathcal{F}_c(c)\Pi$ of the form

$$Q(c)^T \mathcal{F}_c(c)\Pi = \left[\begin{array}{cc} R(c) & S(c) \\ 0 & \Delta(c) \end{array} \right].$$

This can be done in such a way that Q, R, S, and Δ depend smoothly on c. Moreover, $R(c)$ will still be nonsingular and $\Delta(c)$ will be small if we are sufficiently close to the solution set. We then define

$$\mathcal{F}_c(c)^- = \Pi \left[\begin{array}{cc} R(c) & S(c) \\ 0 & 0 \end{array} \right]^+ Q(c)^T,$$

where the superscript $^+$ denotes the Moore-Penrose pseudoinverse, see e. g. [7]. By construction, $\mathcal{F}_c(c)^-$ is an outer inverse of $\mathcal{F}_c(c)$ and depends smoothly on c. One can now show that for this Gauß-Newton process (and sufficiently good initial guess c_0), the assumptions of Theorem 4 in [13] are satisfied giving the claimed superlinear convergence.

6 Conclusions and outlook

In this paper we have presented the theoretical analysis for general over- and underdetermined nonlinear differential-algebraic equations. Such equations include control problems and allow the analysis of systems with redundant equations. We have extended the concept of strangeness index to such general systems and have shown how one can construct a reduced order strangeness-free system, which forms the basis for numerical methods. We have shown that the same approach allows to analyse control problems and we have shown how regularizing state and output feedbacks can be constructed. We have presented the framework of numerical methods to perform these tasks.

References

[1] K. Brenan, S. Campbell, and L. Petzold. *Numerical Solution of Initial-Value Problems in Differential Algebraic Equations.* Elsevier, North Holland, New York, NY, 1989.

[2] K. Brenan, S. Campbell, and L. Petzold. *Numerical Solution of Initial-Value Problems in Differential Algebraic Equations.* Elsevier, North Holland, New York, NY, second edition, 1996.

[3] A. Bunse-Gerstner, V. Mehrmann, and N.K. Nichols. Regularization of descriptor systems by derivative and proportional state feedback. *SIAM Journal Matrix Analysis and its Applications*, 13:46–67, 1992.

[4] A. Bunse-Gerstner, V. Mehrmann, and N.K. Nichols. Regularization of descriptor systems by output feedback. *IEEE Transactions on Automatic Control*, 39:1742–1748, 1994.

[5] R. Byers, T. Geerts, and V. Mehrmann. Descriptor systems without controllability at infinity. *SIAM Journal Control and Optimization*, 35:462–479, 1997.

[6] R. Byers, P. Kunkel, and V. Mehrmann. Regularization of linear descriptor systems with variable coefficients. *SIAM Journal Control and Optimization*, 35:117–133, 1997.

[7] S.L. Campbell and C.D. Meyer. *Generalized Inverses of Linear Tranformations.* Dover Publications, New York, 1979.

[8] S. L. Campbell. A general form for solvable linear time varying singular systems of differential equations. *SIAM Journal Mathematical Analysis*, 18:1101–1114, 1987.

[9] S. L. Campbell. The numerical solution of higher index linear time varying singular systems of differential equations. *SIAM Journal Scientific and Statistical Computing*, 6:334–348, 1988.

[10] S.L. Campbell and E. Griepentrog. Solvability of general differential algebraic equations. *SIAM Journal Scientific Computing*, 16:257–270, 1995.

[11] S.L. Campbell, N.K. Nichols, and W.J. Terrell. Duality, observability, and controllability for linear time-varying descriptor systems. *Circuits, Systems and Signal Processing*, 10:455–470, 1991.

[12] S.L. Campbell and W.J. Terrell. Observability for linear time-varying descriptor systems. *SIAM Journal Matrix Analysis and Applications*, 12:484–496, 1991.

[13] P. Deuflhard and G. Heindl. Affine invariant convergence theorems for Newton's method and extensions to related methods. *SIAM Journal Numerical Analysis*, 16:1–10, 1979.

[14] E. Eich-Söllner and C. Führer. Numerical Methods in Multibody Dynamics. B.G. Teubner, Stuttgart, 1998.

[15] H. Elmquist, S.E. Mattson and M. Otter. Modelica–A language for physical system modeling, visualization and interaction. *Proceedings of the 1999 IEEE Symposium on Computer-Aided Control System Design, CACSD'99, Hawaii, Aug. 22-27, 1999*, CD ROM, see also `http://www.modelica.org/papers/papers.shtml`

[16] E. Griepentrog and R. März. *Differential-Algebraic Equations and Their Numerical Treatment.* Teubner Verlag, Leipzig, 1986.

[17] M. Günther and P. Rentrop. Multirate row-methods and latency of electric circuits. Technical Report TUM-M9208, Mathematisches Institut, TU München, München, FRG, 1992.

[18] S. Hagel. Systematische Modellbildung mechanischer Systeme auf der Basis von Teilkomponenten. In P.C. Müller et al, editors, *Notes of the Workshop 'Identifizierung-, Analyse und Entwurfsmethoden für mechanische Mehrkörpersysteme in Deskriptorform' Paderborn 1994*, pages 67–72, 1994.

[19] H. Hahn and R. Wehage. Dynamic simulation of terrain vehicles. In W. Schiehlen, editor, *Multibody Systems Handbook*, pages 491–503. Springer Verlag Berlin, 1990.

[20] E. Hairer and C. Lubich. Invariant tori of dissipatively perturbed Hamiltonian systems under symplectic discretization. *Applied Numerical Mathematics*, 29:57–71, 1999.

[21] E. Hairer, S.P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I.* Springer-Verlag, Berlin, 1987.

[22] P. Kunkel and V. Mehrmann. Canonical forms for linear differential-algebraic equations with variable coefficients. *Journal of Computational and Applied Mathematics*, 56:225–251, 1994.

[23] P. Kunkel and V. Mehrmann. A new look at pencils of matrix valued functions. *Linear Algebra and its Applications*, 212/213:215–248, 1994.

[24] P. Kunkel and V. Mehrmann. Local and global invariants of linear differential-algebraic equation and their relation. *Electronic Transactions on Numerical Analysis*, 4:138–157, 1996.

[25] P. Kunkel and V. Mehrmann. A new class of discretization methods for the solution of linear differential-algebraic equations. *SIAM Journal on Numerical Analysis*, 33:1941–1961, 1996.

[26] P. Kunkel and V. Mehrmann. Regular solutions of nonlinear differential-algebraic equations and their numerical determination. *Numerische Mathematik*, 79:581–600, 1998.

[27] P. Kunkel, V. Mehrmann, and W. Rath. Analysis and numerical solution of control problems in descriptor form. Technical Report SFB393/99-01, Sonderforschungsbereich 393, Technische Universität Chemnitz.

[28] P. Kunkel, V. Mehrmann, W. Rath, and J. Weickert. A new software package for the solution of linear differential algebraic equations. *SIAM Journal Scientific Computing*, 18:115 – 138, 1997.

[29] J.W. Polderman and J.C. Willems. *Introduction to Mathematical Systems Theory: A Behavioural Approach* , Springer Verlag, New York, 1998.

[30] P.J. Rabier and W.C. Rheinboldt. A geometric treatment of implicit differential–algebraic equations. *Journal on Differential Equations*, 109:110–146, 1995.

[31] P.J. Rabier and W.C. Rheinboldt. Classical and generalized solutions of time-dependent linear differential algebraic equations. *Linear Algebra and its Applications*, 1996.

[32] W. Rath. Derivative and proportional state feedback for linear descriptor systems with variable coefficients. *Linear Algebra and its Applications*, 260:273–310, 1997.

[33] W. Rath. *Feedback Design and Regularization for Linear Descriptor Systems with Variable Coefficients.* Dissertation, Fak. f. Mathematik, TU Chemnitz, 1997. Shaker Verlag, Aachen 1997.

[34] B. Simeon, F. Grupp, C. Führer, and P. Rentrop. A nonlinear truck model and its treatment as a multibody system. Technical Report TUM-M9204, Mathematisches Institut, TU München, München, FRG, 1992.

[35] M. Spivak. *A comprehensive Introduction to Differential Geometry*, volume I. Publish or Perish, Wilmington, Del., 1979.