Technische Universität Berlin
Institut für Mathematik

Index reduction for differential-algebraic equations by minimal extension

P. Kunkel and V. Mehrmann

Technical Report 719-01

Preprint-Reihe des Instituts für Mathematik
Technische Universität Berlin

Report 719-01 May 2003

Index reduction for differential-algebraic equations by minimal extension*

Peter Kunkel † Volker Mehrmann ‡

May 7, 2003

Abstract

In this paper a new index reduction technique is discussed for the treatment of differential-algebraic systems for which extra structural information is available. Based on this information reduced derivative arrays are formed and instead of using expensive subspace computations the index reduction is obtained by introducing new variables.

The new approach is demonstrated for several important classes of differential-algebraic systems, where the structural information is available. These include multibody systems and circuit simulation problems.

The effectiveness of the new approach is demonstrated via numerical examples.

Keywords: differential-algebraic equation, index reduction, minimal extension, circuit simulation, multi-body system, strangeness index, regularization

AMS(MOS) subject classification: 65L05, 34H05

1 Introduction

In this paper we study general over- and under-determined 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. (Here $C^k(S, \mathbb{R}^m)$ denotes the k times continuously differentiable functions from a set S to $\mathbb{R}^m$).

For such general systems of differential-algebraic equations recently a new theoretical analysis has been presented with a general existence and uniqueness theory, see [15, 16, 17, 19, 20, 21]. In particular, a general index concept

*Revised version

†Mathematisches Institut, Universität Leipzig, Augustusplatz 10–11, D-04109 Leipzig, Fed. Rep. Germany. Supported by DFG research grant Ku964/4.

‡Institut für Mathematik, MA 4-5, Technische Universität Berlin, Straße des 17. Juni 136, D-10623 Berlin, Fed. Rep. Germany. Supported by DFG research grant Me790/11.

has been introduced, the strangeness index μ, which generalizes other index concepts as given, e. g. in [4, 10, 14, 28] to systems that are over- or under-determined. Based on the new theoretical analysis also new numerical methods have been introduced in [17, 19, 20, 22] that allow to solve general over- and under-determined systems of arbitrary strangeness index.

The basic idea of the general approach is to consider the original system together with a sufficient number of its derivatives (as a derivative array, see [5]) and to derive locally at every integration step a system of strangeness index 0 that has the same solution set as the original system but contains all the information on the manifold in which the dynamics of the system takes place.

In particular, the derived system consists of two parts, a purely algebraic system describing the manifold of constraints and a differential part describing the dynamics on this manifold. Since all constraints are included it is guaranteed that all the algebraic equations that describe the manifold are satisfied up to the accuracy that is used to solve these equations and hence a drift from the solution manifold is avoided. Note that index reduction methods that do not use all constraints of the system or introduce additional parameters work with a larger set of differential equations than the actual dynamics consists of and may therefore change the stability properties of the original problem. Since in our approach we work with the full set of constraints, instabilities as they are described in [24] do not occur.

In the general case of under- or overdetermined problems (1) the derived system of strangeness index 0 cannot be overdetermined since we must assume consistence of the equations. The resulting possibly underdetermined problem can be treated by all integration methods that work for systems of differentiation index 1, combined with special techniques to deal with possible non-uniqueness, see [18, 20, 21]. In principle, this approach provides a uniform framework for the analysis and numerical solution of differential-algebraic systems. But as is common for general approaches, the computational complexity for this new approach is substantial and makes it in general not feasible for medium or large scale problems, even with its modifications for the use on parallel computers, see [3].

The reason for the high computational complexity is that from the derivative array (which is a system of $(\mu + 1)m$ equations, where μ is the strangeness index) certain nullspaces of the Jacobians and associated projectors onto these nullspaces have to be computed at every integration step. This makes the general method impracticable for large scale problems.

Many practical applications, however, lead to systems of equations with a particular structure that is not reflected in the general approach. It is the topic of this paper to study how the knowledge of extra structure can be used to derive methods that are applicable for higher index systems but are also competitive for medium or large scale problems.

The main ideas that we present in this paper rely on structural information about the equations that lead to high index. This extra information is used to create a reduced size derivative array, so that the computational effort per integration step is highly reduced. But even with these improvements, the general technique would still not be competitive for large scale problems. Even for the reduced size derivative array local nullspace computations are required that may be prohibitive due to the large storage requirements and arithmetic complexity. To deal with this difficulty we modify another index reduction concept that was introduced in [25]. The basic idea of this approach is to introduce new variables, so called *dummy derivatives* to reduce the index. In [25] the necessary decisions, which equations to differentiate and which new variables to introduce, is based on the Pantelidis algorithm [27]. This algorithm is a purely combinatorial method and hence well suited for large scale problems, but it has two major disadvantages. First of all it only produces generic results that may be very sensitive in the neighborhood of non-generic points, and secondly it has recently been shown that it can produce wrong results in certain circumstances [30]. But even if these problems would not occur, the approach of [25] may lead to bigger systems than necessary. We will demonstrate this in Section 3 and introduce a modification of the idea of dummy derivatives that we call *index reduction by minimal extension*.

In Section 4 we discuss the specific structures arising in the simulation of electrical circuits [11, 12, 34]. The dimension of these problems is typically very large and only few components contribute to a higher index. For these systems, recently a detailed (mainly combinatorial) analysis of different circuit elements and the network topology and their contribution to higher index has been given in [8, 11, 12, 34].

Using this structural information it is possible to determine those equations (and there are typically only very few) that lead to a higher index. These techniques provide an inexpensive way to analyze specific circuit models and give indicators, where numerical integration methods may have stability problems [11, 12]. We also show how the structural information can be used to perform the index reduction by minimal extension.

Another major class of differential-algebraic systems arises in the simulation of multi-body systems. In this well studied area [7, 32] index reduction based on structural information is well known. The combination of the general methods of [20] with knowledge about the structural properties and their use in industrial simulation packages has recently been discussed in [1]. In Section 5 we will briefly discuss this topic and compare the index reduction by minimal extension with other stabilization techniques in multi-body system dynamics [4, 7, 9].

In Section 6 we demonstrate the effectiveness of the index reduction via minimal extension with some numerical tests.

2 Preliminaries

The concepts for differential-algebraic equations (DAEs) have changed substantially in recent years. For this reason we recall some of the terminology and some of the previous results that are necessary for the understanding of the new approach.

Definition 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

$$x(t_0) = x_0, \tag{2}$$

if x is a solution of (1) and satisfies (2). An initial condition (2) is called *consistent* if the corresponding initial value problem has at least one solution.

As basis for the existence of solutions and the numerical methods, in [19, 20, 21], hypotheses have been formulated that lead to an index concept, the so-called *strangeness index*, which generalizes the concept of the differentiation index [4]. Let us briefly recall this concept and assume for convenience that all functions are sufficiently smooth.

As in [19], we introduce a nonlinear derivative array, see also [5, 6], of the form

$$F_\ell(t, x, \dot{x}, \ldots, x^{(\ell+1)}) = 0, \tag{3}$$

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

$$F_\ell(t, x, \dot{x}, \ldots, x^{(\ell+1)}) = \begin{bmatrix} F(t, x, \dot{x}) \\ \frac{d}{dt} F(t, x, \dot{x}) \\ \vdots \\ \frac{d^\ell}{dt^\ell} F(t, x, \dot{x}) \end{bmatrix}. \tag{4}$$

Here 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.,

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

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

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

$$\mathbb{L}_\mu = \{ z_\mu \in \mathbb{I} \times \mathbb{R}^n \times \mathbb{R}^n \times \ldots \times \mathbb{R}^n \mid F_\mu(z_\mu) = 0 \}. \tag{5}$$

The following hypothesis was introduced in [20], see also [17, 19].

Hypothesis 1 *Consider a 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*

$$\operatorname{rank} F_{\mu;x,\dot{x},\ldots,x^{(\mu+1)}} = r \tag{6}$$

   on $\mathbb{L}_\mu$.

3. *We have*

$$\operatorname{corank} F_{\mu;x,\dot{x},\ldots,x^{(\mu+1)}} - \operatorname{corank} F_{\mu-1;x,\dot{x},\ldots,x^{(\mu)}} = v \tag{7}$$

   on $\mathbb{L}_\mu$, where the corank is the dimension of the corange and $\operatorname{corank} F_{-1;x} = 0$ *by convention.*

4. *We have*
$$\operatorname{rank} F_{\mu;\dot{x},\ldots,x^{(\mu+1)}} = r - a \tag{8}$$
on $\mathbb{L}_\mu$*, such that there are smooth matrix functions* Z_2 *and* T_2 *defined on* $\mathbb{L}_\mu$ *of size* $((\mu+1)m, a)$ *and* $(n, n-a)$*, respectively, having full rank and satisfying*
$$Z_2^T F_{\mu;\dot{x},\ldots,x^{(\mu+1)}} = 0, \quad \operatorname{rank} Z_2^T F_{\mu;x} = a, \quad Z_2^T F_{\mu;x} T_2 = 0 \tag{9}$$
on $\mathbb{L}_\mu$.

5. *We have*
$$\operatorname{rank} F_{\dot{x}} T_2 = d = m - a - v \tag{10}$$
on $\mathbb{L}_\mu$ *such that there is a smooth matrix function* Z_1 *defined on* $\mathbb{L}_\mu$ *of size* (m, d) *with* $Z_1^T F_{\dot{x}} T_2$ *having full rank.*

The smallest possible μ in Hypothesis 1 is called the *strangeness index* of (1). Systems with vanishing strangeness index are called *strangeness-free* and systems with $m = n$ and $v = 0$ are called *regular*.

If F is sufficiently smooth and satisfies Hypothesis 1 with μ, r, a, d, v, then every solution of (1) also solves a reduced problem consisting of d differential and a algebraic equations and under some further assumptions the converse also holds, see [20].

The results in [20] directly lead to methods for the numerical solution of over- or under-determined systems of the form (1). To compute a consistent initial value at time t_0, i. e., a value x_0 that satisfies the algebraic constraints, we must solve
$$F_\mu(t_0, x_0, \dot{x}_0, \ldots, x_0^{(\mu+1)}) = 0 \tag{11}$$
for $(x_0, \dot{x}_0, \ldots, x_0^{(\mu+1)})$. The classical approach to solve such systems is the Gauß-Newton method, see, e. g., [26]. To perform an integration step from t_0 to $t_1 = t_0 + h$, using for example a BDF-discretization method, 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., [4]), we obtain
$$F_\mu(t_1, x_1, \dot{x}_1, \ldots, x_1^{(\mu+1)}) = 0, \tag{12}$$
$$\hat{Z}_1^T F(t_1, x_1, D_h x_1) = 0, \tag{13}$$
where $\hat{Z}_1$ is a fixed approximation to Z_1 introduced in Hypothesis (1). This system is solved for $(x_1, \dot{x}_1, \ldots, x_1^{(\mu+1)})$ using again the Gauß-Newton method. See [20] for more details.

Analyzing this approach we see that the computational effort in each step of this procedure (apart from the necessary function evaluations) has two parts, the determination of the approximation $\hat{Z}_1$ to Z_1 and the solution of the system consisting of (13) and (12).

In contrast to this, a direct substitution of $\dot{x}$ by $D_h x_1$ in (1) seems a lot less expensive, in particular for large scale systems with structure. However, it is well known that for systems of strangeness index larger than 0 (differentiation index larger than 1 if defined) numerical stability problems may arise or this approach may not work at all, see [4, 14, 13, 33].

But if extra information is available, as for example in multi-body system dynamics or circuit simulation, then we should be able to use this extra information to simplify the computationally expensive parts in the general procedure and thus avoid the numerical problems arising in the direct discretization. We will discuss two modifications in this direction.

The first modification is the identification of equations that have to be differentiated and added to the system. By definition the complete derivative array is used to determine Z_1 or a suitable approximation $\hat{Z}_1$, and to perform the next integration step in (12). If, however, the structure of the problem allows to identify the equations that have to be differentiated, then we do not have to work with the complete derivative array but with a (possibly much) smaller system that replaces F_μ in (12). This smaller system is called *reduced derivative array* in the following.

To obtain this reduced derivative array, let Π_j be a (smooth) matrix function of size (p_j, m) with pointwise orthogonal columns such that

$$\Pi_j(t, x, \dot{x})^T F(t, x, \dot{x}) = 0 \tag{14}$$

describes the equations that are responsible for strangeness index j but not for higher strangeness index.

Here the important assumption is that these projectors are easily available due to the special structure of the problem. We discuss this for circuit simulation in Section 4 and for multi-body systems in Section 5.

The reduced derivative array

$$F_\mu^r(t, x, \dot{x}, \ldots, x^{(\mu+1)}) = 0 \tag{15}$$

is given by the original equation (1) together with all equations

$$\frac{d^l}{dt^l}(\Pi_j(t, x, \dot{x})^T F(t, x, \dot{x})) = 0, \quad j = 1, \ldots, \mu, \; l = 1, \ldots, j. \tag{16}$$

While the system that has to be solved in (12) consists of $(\mu + 1)m$ equations, system (15) only consists of $m + p_1 + 2p_2 + \ldots + \mu p_\mu$ equations, which in many applications is much smaller, see Sections 4 and 5.

The reduced derivative array not only allows to reduce the computational effort in the solution of (12) but it also reduces the complexity of computing the projector Z_1 or an approximation to it, since we can replace the determination of Z_2 and T_2 from the Jacobian of F_μ by corresponding computations from the smaller Jacobian of F_μ^r. But even with the reduction of computational work due to a reduced derivative array, the computation of these projectors may still be infeasible for large scale systems. Therefore, we also discuss another modification which avoids the computation of Z_1 by introducing a minimal number of new variables that lead to an index reduction.

We may summarize the two modifications in the general procedure in the following algorithmic framework.

Algorithm (Index reduction by minimal extension)
Consider a system of differential-algebraic equations of the form (1).

1. Identify the equations that are responsible for a strangeness index larger than 0.
2. Differentiate all equations that are responsible for strangeness index j, but not for higher strangeness index, j times and stack all these equations together with the original system to obtain a reduced derivative array.
3. Identify the minimal number of new variables that have to be introduced.
4. Introduce new variables to obtain the minimally extended strangeness-free system.

This algorithmic framework is feasible, in particular for large scale problems, only if the two identification steps 1. and 3. can be performed without large computational effort, i. e., for example if structural information can be used. We will discuss these two identification steps for several general classes of problems in Section 3 and then for circuit simulation and multi-body dynamics in Sections 4 and 5.

3 Index reduction by minimal extension

In the previous section we have discussed that the computation of the projector Z_1 in (13) is a serious computational bottleneck. As an alternative to the removal of equations from (1), we may increase the number of variables by introducing new variables in the reduced derivative array (15), so that this leads to a new system of strangeness index 0. In general such an approach needs about the same computational effort as the computation of Z_1 and then leads to a larger system in each integration step. But with extra information available this approach may become feasible for large scale problems. The idea of introducing new variables to reduce the index is not new, see [2, 25] or stabilization techniques in multi-body system dynamics [7, 9]. Our new approach, however, leads to a minimal extension and is therefore preferable for large scale systems.

As a motivation and to understand the principle of minimal extension, we begin with linear systems with variable coefficients.

3.1 Linear systems in condensed form

To illustrate the procedure of index reduction by minimal extension, let us first consider linear systems with variable coefficients

$$E(t)\dot{x} = A(t)x + f(t), \tag{17}$$

with $E, A \in C(\mathbb{I}, \mathbb{R}^{m,n})$. It was shown in [15] that under some constant rank assumptions there exist nonsingular matrix valued functions $P \in C(\mathbb{I}, \mathbb{R}^{m,m})$

and $Q \in C^1(\mathbb{I}, \mathbb{R}^{n,n})$ such that the transformed system (with $x(t) = Q(t)y(t)$)

$$P(t)E(t)Q(t)\dot{y} = (P(t)A(t)Q(t) - P(t)E(t)\dot{Q}(t))y + P(t)f(t)$$

has the form (without arguments)

$$\begin{bmatrix} I_s & 0 & 0 & 0 \\ 0 & I_d & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \end{bmatrix} = \begin{bmatrix} 0 & A_{12} & 0 & A_{14} \\ 0 & 0 & 0 & A_{24} \\ 0 & 0 & I_a & 0 \\ I_s & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \end{bmatrix}, \quad (18)$$

where the fourth block column has size u and the fifth block row has size v.

It follows immediately from the results in [15] that system (18) has strangeness index 0 if and only if $s = 0$ and it has strangeness index 1 if and only if $\operatorname{rank} A_{14} = \operatorname{rank}[A_{12}\ A_{14}]$. It is then obvious that the equations that have to be differentiated to reduce the index of the system are exactly given by the fourth block row of this system.

Differentiating these equations and adding the derivatives to the system we obtain the reduced derivative array

$$\begin{bmatrix} I_s & 0 & 0 & 0 \\ 0 & I_d & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ I_s & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \end{bmatrix} = \begin{bmatrix} 0 & A_{12} & 0 & A_{14} \\ 0 & 0 & 0 & A_{24} \\ 0 & 0 & I_a & 0 \\ I_s & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \\ \dot{g}_4 \end{bmatrix}. \quad (19)$$

In the general approach [15, 17, 20] we would now compute nullspaces and transformations of the two system matrices of this reduced derivative array and after this remove some equations to obtain a new system with the same solution set as (18). In view of the discussed complexity problems, the idea of minimal extension is to introduce a minimal number (here s) new variables to reduce the index. In this special case we replace every occurrence of $\dot{y}_1$ in (19) by the new variable y_5 and obtain the extended system

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & I_d & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \\ \dot{y}_5 \end{bmatrix} = \begin{bmatrix} 0 & A_{12} & 0 & A_{14} & -I_s \\ 0 & 0 & 0 & A_{24} & 0 \\ 0 & 0 & I_a & 0 & 0 \\ I_s & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -I_s \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \\ \dot{g}_4 \end{bmatrix}. \quad (20)$$

It is easy to see that if $(y_1, \ldots, y_5)$ solves (20) then $(y_1, \ldots, y_4)$ solves (19) and conversely if $(y_1, \ldots, y_4)$ solves (19) then $(y_1, \ldots, y_4, \dot{y}_1)$ solves (20).

The following lemma shows when this system of size $(m + s, n + s)$ has strangeness index 0.

Lemma 2 *The differential-algebraic system in condensed form (18) has strangeness index* 1 *if and only if the extended system (20) has strangeness index* 0*.*

Proof. As in [15] we compute matrices Z, T whose columns span the left and right nullspace, respectively, of the coefficient of $\dot{y}$. A possible choice is

$$Z = \begin{bmatrix} I_s & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & I_a & 0 & 0 & 0 \\ 0 & 0 & I_s & 0 & 0 \\ 0 & 0 & 0 & I_v & 0 \\ 0 & 0 & 0 & 0 & I_s \end{bmatrix}, \quad T = \begin{bmatrix} I_s & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & I_a & 0 & 0 \\ 0 & 0 & I_u & 0 \\ 0 & 0 & 0 & I_s \end{bmatrix}.$$

Then let the columns of

$$V = \begin{bmatrix} V_{11} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & I_u \\ -V_{11} & 0 \end{bmatrix}$$

with $V_{11}^T A_{14} = 0$ span the left nullspace of $Z^T AT$ and let

$$T' = \begin{bmatrix} 0 \\ I_d \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

complete T to a nonsingular matrix.

It follows that $V^T Z^T AT' = 0$ if and only if $V_{11}^T A_{12} = 0$, which is equivalent to $\operatorname{rank}(A_{14}) = \operatorname{rank}([A_{12}\ A_{14}])$, i. e., that the system has strangeness index 1. □

We have seen that the introduction of new variables reduces the index of the system. Since from the theory of [15, 17] there are at least s couplings between algebraic and differential equations which must be removed, any smaller extension would still have at least one coupling and hence can not be strangeness-free. Thus, we have the following lemma.

Lemma 3 *Suppose that the differential-algebraic system in condensed form (18) has strangeness index* 1*. Then the minimal number of new variables that have to be introduced in the reduced derivative array (19) so that the extended system has strangeness index* 0 *is* s*.*

Remark 1 At this point we can already formulate a general principle how to obtain a minimally extended system from a given DAE. Having detected a variable, say w, in a purely algebraic condition that also occurs in differentiated form (as $\dot{w}$) in a different part of the system (as, e. g., y_1 in (18)) it is clear that this coupling contributes to a higher index. Eliminating the crucial variable (namely $\dot{w}$) with the help of the differentiated algebraic condition can be seen as the basic step towards a reduced problem that has the same size (and the same solution set) as the original problem. The basic step for the construction of a corresponding minimally extended system on the other side is given by adding the differentiated algebraic condition to the given problem and replacing the crucial variable (still $\dot{w}$) by a new one (setting, e. g., $z = \dot{w}$). Since elimination of the new variable would simply result in the same system as for the basic step towards a reduced problem, the solutions of the extended system only differs in the additional component due to the introduction of a new variable. This general principle will also show up in the following sections, even when we treat nonlinear problems. The critical step will only be to identify the crucial variables.

As we have already discussed in the introduction, the concept of introducing new (dummy) variables to reduce the index of a system is not new, it has originally been introduced in [25] although in a different and not necessarily minimal way. To see this let us study the following example from [25].

Example 1 Consider the second order differential-algebraic system

$$\begin{aligned} x_1 + x_2 + u_1(t) &= 0, \\ x_1 + x_2 + x_3 + u_2(t) &= 0, \\ \dot{x}_3 + x_1 + x_4 + u_3(t) &= 0, \\ 2\ddot{x}_1 + \ddot{x}_2 + \ddot{x}_3 + \dot{x}_4 + u_4(t) &= 0, \end{aligned}$$

where u is a given forcing function. If we write this system as first order system with $x_5 = \dot{x}_1$, $x_6 = \dot{x}_2$ and $x_7 = \dot{x}_3$, we obtain the system

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \dot{x}_6 \\ \dot{x}_7 \end{bmatrix} =$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ u_1 \\ u_2 \\ -u_3 \\ -u_4 \end{bmatrix}.$$

Computing the form (18) for this system we obtain $d = 2,\ a = 1,\ s = 2$ and the strangeness index $\mu = 1$. It follows that the minimally extended system is a first order system of 9 equations. In contrast to this, the system obtained by introducing dummy derivatives as in [25] is a second order system of 9 equations, which can be rewritten as a first order system of 10 equations, since only one variable occurs with second derivative.

A similar index reduction procedure by minimal extension can also be applied to systems in normal form that have strangeness index higher than 1. Since this approach becomes very technical, see also [15], we present here only the case of a uniquely solvable system in normal form (18) that has no redundant equations and strangeness index $\mu = 2$. Permute this form by exchanging the first two block rows and columns to

$$\begin{bmatrix} I_{d_0} & 0 & 0 & 0 \\ 0 & I_{s_0} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & A_{14} \\ A_{21} & 0 & 0 & A_{24} \\ 0 & 0 & I_{a_0} & 0 \\ 0 & I_{s_0} & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix}. \quad (21)$$

Then determine (locally) permutations Π, Ψ such that

$$\Pi^T A_{24} \Psi = \begin{bmatrix} \tilde{A}_{25} & \tilde{A}_{26} \\ \tilde{A}_{35} & \tilde{A}_{36} \end{bmatrix}$$

and $\tilde{A}_{25}$ is invertible with $\operatorname{rank} \tilde{A}_{25} = \operatorname{rank} A_{24}$. Multiplying the second and last block row from the left by

$$V_2 = \begin{bmatrix} I & 0 \\ -\tilde{A}_{35}\tilde{A}_{25}^{-1} & I \end{bmatrix} \Pi^T$$

and making the change of variables

$$\begin{bmatrix} z_2 \\ z_3 \end{bmatrix} = V_2 y_2, \quad z_5 = \Psi^T \begin{bmatrix} I & \tilde{A}_{25}^{-1}\tilde{A}_{26} \\ 0 & I \end{bmatrix} y_4$$

we obtain a transformed system

$$\begin{bmatrix} I_{d_0} & 0 & 0 & 0 & 0 & 0 \\ 0 & I_{s_{0,1}} & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{s_{0,2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \\ \dot{z}_4 \\ \dot{z}_5 \\ \dot{z}_6 \end{bmatrix} =$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 & \tilde{A}_{15} & \tilde{A}_{16} \\ \tilde{A}_{21} & 0 & 0 & 0 & \tilde{A}_{25} & 0 \\ \tilde{A}_{31} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{a_0} & 0 & 0 \\ 0 & I_{s_{0,1}} & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{s_{0,2}} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \end{bmatrix} + \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \end{bmatrix}. \quad (22)$$

This system has strangeness index $\mu = 2$ if and only if $s_0 \neq 0$, the matrix functions $\tilde{A}_{31}$ and $\tilde{A}_{16}^T$ have full row rank $s_1 = s_{0,1}$ and $\tilde{A}_{31}\tilde{A}_{16}$ is nonsingular.

Note that we have assumed that there are no redundant equations and that the system is uniquely solvable. Hence, there (locally) exists a permutation matrix Π_1 such that the leftmost columns C_1 of $C = [\, C_1 \; C_2 \,] = \tilde{A}_{31}\Pi_1$ form a nonsingular matrix function of size (s_1, s_1).

This information is sufficient to determine the equations that have to be differentiated and the variables that have to be replaced. Partitioning

$$z_1\Pi_1 = \begin{bmatrix} z_{11} \\ z_{21} \end{bmatrix},$$

adding the equations

$$\dot{z}_2 = -\dot{h}_5,\ \dot{z}_3 = -\dot{h}_6,\ 0 = C\Pi_1^T\dot{z}_1 + \dot{C}\Pi_1^T z_1 + \dot{h}_3 + \ddot{h}_6$$

and introducing the new variables $z_7 = \dot{z}_2$, $z_8 = \dot{z}_3$, $z_9 = \dot{z}_{1,1}$, we obtain a system of strangeness index 0. Note that this is again a minimal extension, since we have added $s_0 + s_1$ equations and variables and it follows from the theory in [15] that in this case this is the number of couplings between differential and algebraic equations that have to be removed.

Looking in detail at the index reduction process described in [15] it follows how to proceed if the strangeness index is larger than 2 or if the system is not uniquely solvable.

In this section we have seen that for systems in condensed form (18) we can avoid the computational effort of transforming the system into an equivalent system of the same size with lower index if we introduce a minimal number of new variables. But typically a system is not in the condensed from (18) and also it is computationally expensive to compute this form [17], in particular, for large scale systems.

3.2 Linear semi-explicit systems

If the system under consideration is linear and semi-explicit then the index reduction by minimal extension is also easily performed if the algebraic equations that lead to the higher index can be identified.

Consider the regular linear semi-explicit system $E(t)\dot{x} = A(t)x + f$ of $m = m_1 + m_2$ equations in m unknowns

$$\begin{bmatrix} I_{m_1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11}(t) & A_{12}(t) \\ A_{21}(t) & A_{22}(t) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix} \tag{23}$$

with sufficiently smooth matrix functions A_{ij} and strangeness index $\mu = 1$. Suppose further that we can identify the equations that lead to a strangeness index larger than 0. If the computational effort is feasible then this can for example be done by determining the rank a of A_{22} and by determining (locally) permutation matrices Π and Ψ such that, with the change of variables $y = \Psi^T x$, the system

$$\Pi^T E\Psi\dot{y} = \Pi^T A(t)\Psi y + \Pi f$$

can be written as (renaming the blocks in A and leaving off arguments)

$$\begin{bmatrix} I_s & 0 & 0 & 0 \\ 0 & I_d & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix}, \quad (24)$$

where the matrix functions A_{33} of size (a, a) and $\tilde{A}_{41} = A_{41} - A_{43}A_{33}^{-1}A_{31}$ of size (s, s) are nonsingular. If we multiply the system by

$$\begin{bmatrix} I_s & 0 & 0 & 0 \\ 0 & I_d & 0 & 0 \\ 0 & 0 & I_a & 0 \\ 0 & 0 & -A_{43}A_{33}^{-1} & I_s \end{bmatrix}$$

from the left then the resulting system is

$$\begin{bmatrix} I_s & 0 & 0 & 0 \\ 0 & I_d & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ \tilde{A}_{41} & \tilde{A}_{42} & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ \tilde{g}_4 \end{bmatrix}, \quad (25)$$

with $\tilde{A}_{42} = A_{42} - A_{43}A_{33}^{-1}A_{32}$ and $\tilde{g}_4 = g_4 - A_{43}A_{33}^{-1}g_3$. Note that the diagonal block $\tilde{A}_{44}$ vanishes by construction of A_{33}.

We then add the derivative of the last block row of (25) to the system and introduce new variables for $\dot{y}_1$, i. e., in the permuted system (24) we set $\dot{y}_1 = y_5$ and obtain the minimally extended system

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & I_d & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \tilde{A}_{42} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \\ \dot{y}_5 \end{bmatrix} =$$

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & -I_s \\ A_{21} & A_{22} & A_{23} & A_{24} & 0 \\ A_{31} & A_{32} & A_{33} & A_{34} & 0 \\ A_{41} & A_{42} & A_{43} & A_{44} & 0 \\ -\frac{d}{dt}\tilde{A}_{41} & -\frac{d}{dt}\tilde{A}_{42} & 0 & 0 & -\tilde{A}_{41} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ -\frac{d}{dt}\tilde{g}_4 \end{bmatrix}. \quad (26)$$

Note that neither the permutation nor the block elimination have to be carried out, they are just needed to identify the equations that have to be differentiated and the variables that have to be replaced. Again for large scale problems such a procedure would be feasible only if s is much smaller than m and the computation of $\tilde{A}_{41}$, $\tilde{A}_{42}$ and $\tilde{g}_4$ can be carried out easily.

We have the following result.

Lemma 4 *The semi-explicit system (23) has strangeness index* 1 *if and only the extended system (26) has strangeness index* 0*.*

Proof. Eliminating $\tilde{A}_{42}$ in the last row on the left hand side of system (26) by a block elimination from the left, we see that (26) has strangeness index 0 if and only if the matrix

$$B = \begin{bmatrix} A_{11} & A_{13} & A_{14} & -I_s \\ A_{31} & A_{33} & A_{34} & 0 \\ A_{41} & A_{43} & A_{44} & 0 \\ -\frac{d}{dt}(\tilde{A}_{41}) - A_{42}A_{21} & -A_{42}A_{23} & A_{42}A_{24} & -\tilde{A}_{41} \end{bmatrix}$$

is nonsingular. Recall that A_{33} and $\tilde{A}_{41}$ are square nonsingular, so B is nonsingular if and only if the matrix

$$C = \begin{bmatrix} A_{13} + \tilde{A}_{41}^{-1}A_{42}A_{23} & A_{14} + \tilde{A}_{41}^{-1}A_{42}A_{24} \\ A_{33} & A_{34} \end{bmatrix}$$

is nonsingular. Considering the transformed system (25) we see that C is nonsingular if and only if (25) and hence the original system has strangeness index 1. $\square$

3.3 Nonlinear semi-explicit systems

In this subsection we consider uniquely solvable nonlinear semi-explicit systems of strangeness index $\mu = 1$ of the form

$$\begin{aligned} \dot{x} &= f(x, y, z), \\ 0 &= g(x, y), \qquad (27) \\ 0 &= h(x), \qquad (28) \end{aligned}$$

with m_1, m_2, m_3 rows in f, g, h, respectively. We assume that the Jacobians g_y and $h_x f_z$ are invertible, which means that the equations $h(x) = 0$ represent the equations that lead to a strangeness index $\mu = 1$.

In this situation the reduced derivative array is

$$\begin{aligned} \dot{x} &= f(x, y, z), \\ 0 &= g(x, y), \\ 0 &= h(x), \\ 0 &= h_x(x)\dot{x}. \qquad (29) \end{aligned}$$

To perform the minimal extension, select m_3 variables x_2 of x such that the partial derivative h_{x_2} is (locally) invertible and split the fourth equation in (29) as

$$h_{x_1}(x_1, x_2)\dot{x}_1 + h_{x_2}(x_1, x_2)\dot{x}_2 = 0. \qquad (30)$$

Replacing every occurrence of the variables x_2 by the new variables $w = \dot{x}_2$, we have rewritten system (29) as

$$\dot{x}_1 = f_1(x_1, x_2, y, z),$$

$$
\begin{aligned}
0 &= f_2(x_1,x_2,y,z) - w,\\
0 &= g(x_1,x_2,y),\\
0 &= h(x_1,x_2),\\
h_{x_1}(x_1,x_2)\dot{x}_1 &= -h_{x_2}(x_1,x_2)w.
\end{aligned} \tag{31}
$$

Lemma 5 *The semi-explicit system (28) has strangeness index $\mu = 1$ if and only the extended system (31) has strangeness index $\mu = 0$.*

Proof. In the present case the characteristic values r, a, d, and v of Hypothesis 1 do not depend on the point where we linearize. Thus, we may ignore the condition $\mathbb{L}_\mu \neq \emptyset$ of Hypothesis 1.

Elimination of the derivative $\dot{x}_1$ in system (31) via the first block row yields the equivalent system

$$
\begin{aligned}
\dot{x}_1 &= f_1(x_1,x_2,y,z),\\
0 &= f_2(x_1,x_2,y,z) - w,\\
0 &= g(x_1,x_2,y),\\
0 &= h(x_1,x_2),\\
0 &= -h_{x_2}(x_1,x_2)w - h_{x_1}(x_1,x_2)f_1(x_1,x_2,y,z).
\end{aligned} \tag{32}
$$

This system has strangeness index 0 if and only if the matrix (without arguments)

$$
\begin{bmatrix}
f_{2;x_2} & f_{2;y} & f_{2;z} & -I\\
g_{x_2} & g_y & 0 & 0\\
h_{x_2} & 0 & 0 & 0\\
-h_{x_1}f_{1;x_2} & -h_{x_1}f_{1;y} & -h_{x_1}f_{1;z} & h_{x_2}
\end{bmatrix} \tag{33}
$$

is invertible, which is the case if and only if the matrices g_y, h_{x_2} and

$$
\begin{bmatrix}
f_{2;z} & -I\\
-h_{x_1}f_{1;z} & -h_{x_2}
\end{bmatrix}
$$

are invertible. This is exactly the condition for the original system to have strangeness index $\mu = 1$. $\square$

3.4 Hessenberg Systems

Another important class of systems for which we obtain reduced derivative arrays and a minimal extension without much computational effort are Hessenberg systems, see [4]. In the linear case, a Hessenberg system has the form

$$
\begin{bmatrix}
I & 0 & 0 & \dots & 0\\
0 & I & 0 & \dots & 0\\
0 & 0 & \ddots & & \vdots\\
\vdots & \dots & \dots & I & 0\\
0 & \dots & \dots & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\dot{x}_1\\ \dot{x}_2\\ \vdots\\ \dot{x}_{r-1}\\ \dot{x}_r
\end{bmatrix} =
$$

$$\begin{bmatrix} B_{11} & B_{12} & \dots & \dots & B_{1,r} \\ B_{21} & B_{22} & \dots & B_{2,r-1} & 0 \\ 0 & B_{32} & \ddots & B_{3,r-1} & 0 \\ \vdots & 0 & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & B_{r,r-1} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{r-1} \\ x_r \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{r-1} \\ f_r \end{bmatrix}, \qquad (34)$$

with $B_{r,r-1}B_{r-1,r-2}\cdots B_{2,1}B_{1,r}$ nonsingular.

Hessenberg systems are uniquely solvable and have differentiation index r, see [4], and hence according to [17] for $r > 0$ strangeness index $r-1$.

For $r = 2$ the Hessenberg system (34) of size (m, m) has the form

$$\begin{bmatrix} I_{m_1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \qquad (35)$$

with $B_{21}B_{12}$ invertible. Here, from the structure, we see that a transformation to the normal form (18) would lead to $s = m - m_1$. Let Π_1 be a permutation matrix such that (locally) the first s columns of $B_{21}\Pi_1$ form a nonsingular matrix. Multiplying the first row of (35) by Π_1^T, setting $y_3 = x_2$, and performing the change of variables

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \Pi^T x_1$$

with y_1 of size s we obtain a system of the form

$$\begin{bmatrix} I_s & 0 & 0 \\ 0 & I_{m_1-s} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}, \qquad (36)$$

with A_{31} nonsingular. This system has strangeness index $\mu = 1$ if and only if $A_{31}A_{13} + A_{32}A_{23}$ is nonsingular.

In this way we have identified the last block row of (36) as the equations that have to be differentiated to obtain the reduced derivative array and that we should introduce the new variables $y_4 = \dot{y}_1$. We obtain the minimally extended strangeness-free system

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & I_{m_1-s} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & A_{32} & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \end{bmatrix} =$$

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} & -I_s \\ A_{21} & A_{22} & A_{23} & 0 \\ A_{31} & A_{32} & 0 & 0 \\ -\dot{A}_{31} & -\dot{A}_{32} & 0 & -A_{31} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ -\dot{g}_3 \end{bmatrix}. \qquad (37)$$

That this system has strangeness index 0 follows directly from the nonsingularity of the matrix

$$\begin{bmatrix} A_{11} & A_{13} & -I_s \\ A_{31} & 0 & 0 \\ -\dot{A}_{31} - A_{32}A_{21} & -A_{32}A_{23} & -A_{31} \end{bmatrix}.$$

By this construction we see that for Hessenberg systems of strangeness index 1 the equations that have to be differentiated are clear and to identify the new variables we only have to identify linear independent columns in the block B_{21}.

For $r = 3$ a linear Hessenberg system has the form

$$\begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & I_{m_2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & 0 \\ 0 & B_{32} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}, \qquad (38)$$

with $B_{32}B_{21}B_{13}$ nonsingular. In this case we (locally) determine permutation matrices Π_1, Π_2 such that in

$$B_{32}\Pi_2 = \begin{bmatrix} A_{53} & A_{54} \end{bmatrix}, \quad \Pi_1^T B_{13} = \begin{bmatrix} A_{15} \\ A_{25} \end{bmatrix}, \quad B_{32}B_{21}\Pi_1 = \begin{bmatrix} \hat{A}_{31} & \hat{A}_{32} \end{bmatrix}$$

the matrices A_{53} and $\hat{A}_{31}$ are square and nonsingular. Multiplying the first block row of (38) by Π_1^T and second block row by Π_2^T, setting $y_5 = x_3$, and performing the change of variables

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \Pi_1^T x_1, \quad \begin{bmatrix} y_3 \\ y_4 \end{bmatrix} = \Pi_2^T x_2$$

with y_1 of size s_1 and y_3 of size s_2 we obtain the transformed system

$$\left[\begin{array}{cc|cc|c} I_{s_1} & 0 & 0 & 0 & 0 \\ 0 & I_{d_1} & 0 & 0 & 0 \\ \hline 0 & 0 & I_{s_2} & 0 & 0 \\ 0 & 0 & 0 & I_{d_2} & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} \dot{y}_1 \\ \dot{y}_2 \\ \hline \dot{y}_3 \\ \dot{y}_4 \\ \hline \dot{y}_5 \end{array}\right] =$$

$$\left[\begin{array}{cc|cc|c} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} \\ A_{21} & A_{22} & A_{23} & A_{24} & A_{25} \\ \hline A_{31} & A_{32} & A_{33} & A_{34} & 0 \\ A_{41} & A_{42} & A_{43} & A_{44} & 0 \\ \hline 0 & 0 & A_{53} & A_{54} & 0 \end{array}\right] \left[\begin{array}{c} y_1 \\ y_2 \\ \hline y_3 \\ y_4 \\ \hline y_5 \end{array}\right] + \left[\begin{array}{c} g_1 \\ g_2 \\ \hline g_3 \\ g_4 \\ \hline g_5 \end{array}\right]. \qquad (39)$$

In this system we have that the matrix

$$\begin{aligned} C &= B_{32}B_{21}B_{13} = (B_{32}\Pi_2)(\Pi_2^T B_{21}\Pi_1)(\Pi_1^T B_{13}) \\ &= \begin{bmatrix} A_{53} & A_{54} \end{bmatrix} \begin{bmatrix} A_{31} & A_{32} \\ A_{41} & A_{42} \end{bmatrix} \begin{bmatrix} A_{15} \\ A_{25} \end{bmatrix} \\ &= A_{53}A_{31}A_{15} + A_{54}A_{41}A_{15} + A_{53}A_{32}A_{25} + A_{54}A_{42}A_{25} \end{aligned}$$

is nonsingular. We differentiate the equation

$$A_{53}y_3 + A_{54}y_4 + g_5 = 0$$

and insert the third and fourth equation of (39) to get

$$\hat{A}_{31}y_1 + \hat{A}_{32}y_2 + (\hat{A}_{33} + \dot{A}_{53})y_3 + (\hat{A}_{34} + \dot{A}_{54})y_4 + \hat{g}_3 = 0,$$

where $\hat{A}_{3,i} = A_{5,3}A_{3,i} + A_{5,4}A_{4,i}$ for $i = 1, 2, 3, 4$ and $\hat{g}_3 = A_{53}g_3 + A_{54}g_4 + \dot{g}_5$. Then we add the derivatives of these two equations to the system and introduce new variables $y_6 = \dot{y}_3,\ y_7 = \dot{y}_1$. In this way we obtain the minimally extended system

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_{d_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{d_2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -A_{54} & 0 & 0 & 0 \\ 0 & -\hat{A}_{32} & 0 & -\hat{A}_{34} - \dot{\hat{A}}_{54} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \\ \dot{y}_5 \\ \dot{y}_6 \\ \dot{y}_7 \end{bmatrix} =$$

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & 0 & -I_{s_1} \\ A_{21} & A_{22} & A_{23} & A_{24} & A_{25} & 0 & 0 \\ A_{31} & A_{32} & A_{33} & A_{34} & 0 & -I_{s_2} & 0 \\ A_{41} & A_{42} & A_{43} & A_{44} & 0 & 0 & 0 \\ 0 & 0 & A_{53} & A_{54} & 0 & 0 & 0 \\ 0 & 0 & \dot{A}_{53} & \dot{A}_{54} & 0 & A_{53} & 0 \\ \frac{d}{dt}\hat{A}_{31} & \frac{d}{dt}\hat{A}_{32} & * & * & 0 & \hat{A}_{33} + \dot{A}_{53} & \hat{A}_{31} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \\ \dot{g}_5 \\ \frac{d}{dt}\hat{g}_3 \end{bmatrix}.$$

Here we denote by $*$ a block in the matrix that is not relevant for the index of the system.

This system has strangeness index 0, since, if we eliminate the last two rows in the coefficient of $\dot{y}$ by multiplying the system from the left by the matrix

$$\begin{bmatrix} I_{s_1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_{d_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{s_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{d_2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{s_2} & 0 & 0 \\ 0 & 0 & 0 & A_{54} & 0 & I_{s_2} & 0 \\ 0 & \hat{A}_{32} & 0 & \hat{A}_{34} + \dot{\hat{A}}_{54} & 0 & 0 & I_{s_1} \end{bmatrix},$$

then in the coefficient of y we obtain the relevant submatrix

$$S = \begin{bmatrix} A_{11} & A_{13} & A_{15} & 0 & -I_{s_1} \\ A_{31} & A_{33} & 0 & -I_{s_2} & 0 \\ 0 & A_{53} & 0 & 0 & 0 \\ A_{54}A_{41} & * & 0 & A_{53} & 0 \\ * & * & \hat{A}_{32}A_{25} & * & \hat{A}_{31} \end{bmatrix},$$

obtained by projecting from the left and right by matrices whose columns span the left and right nullspace of the coefficient of $\dot{y}$. Using the identity in the second row to eliminate the other elements and, since A_{53} is nonsingular, the nonsingularity of S is equivalent to the nonsingularity of

$$\begin{bmatrix} A_{11} & A_{15} & -I_{s_1} \\ \hat{A}_{31} & 0 & 0 \\ * & \hat{A}_{32}A_{25} & \hat{A}_{31} \end{bmatrix}$$

and this is equivalent to the nonsingularity of $\hat{A}_{31}$ and C.

How to proceed for Hessenberg systems with $r > 3$ is then canonical. For nonlinear Hessenberg systems the analysis proceeds analogously.

4 Application to circuit simulation

In this section we discuss the application of reduced derivative arrays and minimal extension in the context of circuit simulation. In several recent papers, a detailed analysis has been given, which influence specific elements and their combination have on the index, for a survey see [11, 12]. Furthermore in [8, 34] topological methods have been derived to analyze from the network topology which equations are responsible for higher index and projectors are determined (in a purely combinatorial way) to filter out these equations from the system. We will briefly review these results here, so that we can produce the reduced derivative array. After this we discuss the identification which new variables have to be introduced for the circuit simulation applications. We discuss, in particular, the modified nodal analysis and the charge oriented modified nodal analysis. Denoting by e the node potentials, by j_L and j_V the currents through inductances and voltage sources, respectively, by i and v the functions describing the current and voltage sources, respectively, by r the function describing the resistances, and finally by q_C and ϕ_L the functions describing the charges of the capacitances and the fluxes of the inductances, respectively, one obtains from the modified nodal analysis (MNA), see, e. g., [11], a quasi-linear system of differential-algebraic equations of the form

$$
\begin{aligned}
0 &= A_C \frac{dq_C(A_C^T e, t)}{dt} + A_R r(A_R^T e, t) + A_L j_L + \\
&\qquad A_V j_V + A_I i(A^T e, \frac{dq(A_C^T e, t)}{dt}, j_L, j_V, t), \\
0 &= \frac{d\phi_L(j_L, t)}{dt} - A_L^T e, \\
0 &= A_V^T e - v(A^T e, \frac{dq_C(A_C^T e, t)}{dt}, j_L, j_V, t),
\end{aligned} \tag{40}
$$

where the incidence matrix A containing the information on the topology of the circuit is split as $\left[\, A_C \; A_L \; A_R \; A_V \; A_I \,\right]$, with A_C, A_L, A_R, A_V and A_I describing the branch current relation for capacitive, inductive, resistive branches and branches for voltage sources and current sources, respectively.

For the conventional MNA the vector of unknown variables consists of all node potentials e and all branch currents j_L, j_V of current-controlled elements. Introducing new functions

$$
C(u,t) = \frac{\partial q_C(u,t)}{\partial u}, \quad L(j,t) = \frac{\partial \phi_L(j,t)}{\partial j},
$$

and the notation

$$
q_t(u,t) = \frac{\partial q(u,t)}{\partial t}, \quad \phi_t(j,t) = \frac{\partial \phi(j,t)}{\partial t},
$$

the system is reformulated as

$$
\begin{aligned}
0 &= A_C C(A_C^T e, t) A_C^T \frac{de}{dt} + A_C q_t(A_C^T e, t) + A_R r(A_R^T e, t) + A_L j_L + \\
&\qquad A_V j_V + A_I i(A^T e, C(A_C^T e, t) A_C^T \frac{de}{dt} + A_C q_t(A_C^T e, t), j_L, j_V, t), \\
0 &= L(j_L, t)\frac{j_L}{dt} + \phi_t(j_L, t) - A_L^T e, \\
0 &= A_V^T e - v(A^T e, C(A_C^T e, t) A_C^T \frac{de}{dt} + A_C q_t(A_C^T e, t), j_L, j_V, t).
\end{aligned} \tag{41}
$$

In the charge oriented MNA the vector of unknowns is extended by the charges q of capacitances and the fluxes ϕ of inductances, and the original voltage-charge and current-flux equations are included in the system yielding

$$
\begin{aligned}
0 &= A_C \frac{dq}{dt} + A_R r(A_R^T e, t) + A_L j_L + A_V j_V + A_I i(A^T e, \frac{dq}{dt}, j_L, j_V, t), \\
0 &= \frac{d\phi}{dt} - A_L^T e, \\
0 &= A_V^T e - v(A^T e, \frac{dq}{dt}, j_L, j_V, t), \\
0 &= q - q_C(A_C^T e, t), \\
0 &= \phi - \phi_L(j_L, t).
\end{aligned} \tag{42}
$$

In [8, 11, 12, 34] a detailed analysis of the differentiation index and other properties of these systems have been given. In particular, it has been shown how a purely topological analysis can be used to determined the higher index equations. In this way it is possible to derive the reduced derivative array without extra computational effort. Furthermore, also the identification of the minimal extension can be obtained from this information as we will discuss now.

In [8] the following projectors were introduced. The projector onto $\operatorname{kernel} A_C^T$ was denoted by Q_C, that onto $\operatorname{kernel} A_V^T Q_C$ by Q_{V-C}, that onto $\operatorname{kernel} A_R^T Q_C Q_{VC}$ by Q_{R-CV}, that onto $\operatorname{kernel} Q_C^T A_V$ by $\bar{Q}_{V-C}$ and, furthermore, the product of these projectors was denoted by $Q_{CRV} = Q_C Q_{V-C} Q_{R-CV}$. In abuse of the notation in [8] we use the same terms to denote the full-rank parts of these projectors, i.e., to denote projection matrices whose columns span the corresponding spaces. In this way in the following equations we avoid unnecessary equations of the form $0 = 0$. These constant projection matrices can be obtained by purely topological analysis of the network at essentially no computational cost.

Then for the conventional MNA (41) the equations that are responsible for a strangeness index higher than 0 are given by the projected equations (in the following we omit the arguments of the functions i and v to simplify the notation, see [8] for a detailed derivation of the exact form of these equations)

$$
\begin{aligned}
0 &= Q_{CRV}^T (A_L j_L + A_I i(\,\cdot\,)), \\
0 &= \bar{Q}_{V-C}(A_V^T e - v(\,\cdot\,)).
\end{aligned} \tag{43}
$$

It follows that the reduced derivative array consists of the equations in (41) together with the derivatives of (43)

$$\begin{aligned} 0 &= Q_{CRV}^T(A_L \frac{dj_L}{dt} + A_I \frac{di(\,\cdot\,)}{dt}), \\ 0 &= \bar{Q}_{V-C}(A_V^T \frac{de}{dt} - \frac{dv(\,\cdot\,)}{dt}). \end{aligned} \tag{44}$$

To determine the minimal extension we have to find nonsingular matrices Π_e, Π_j such that

$$Q_{CRV}^T A_L \Pi_j^{-1} = [J_1 \ 0], \quad \bar{Q}_{V-C} A_V^T \Pi_e^{-1} = [F_1 \ 0]$$

with J_1, F_1 square nonsingular. These can be obtained with very small computational effort, since $Q_{CRV}^T A_L$ and $\bar{Q}_{V-C} A_V^T$ are still only incidence-like matrices (containing topological information on the circuit in form of integers) the computation of Π_j, Π_e and their inverses is possible with very small computational effort and very accurately. We partition

$$\tilde{j}_L = \Pi_j j_L = \begin{bmatrix} \tilde{j}_{L_1} \\ \tilde{j}_{L_2} \end{bmatrix}, \quad \tilde{e}_L = \Pi_e e = \begin{bmatrix} \tilde{e}_1 \\ \tilde{e}_2 \end{bmatrix}$$

conformally and introduce new variables

$$\hat{e}_1 = \frac{d\tilde{e}_1}{dt}, \quad \hat{j}_1 = \frac{d\tilde{j}_{L_1}}{dt}. \tag{45}$$

Note that since we add exactly as many equations as needed, the extension is minimal. We have the following result.

Theorem 6 *Let the assumptions of Theorem 2.1 in [8] hold. Then the minimally extended system for the conventional MNA given by the system*

$$\begin{aligned} 0 &= A_C C(A_C^T \Pi_e^{-1} \tilde{e}, t) A_C^T \Pi_e^{-1} \begin{bmatrix} \hat{e}_1 \\ \frac{d\tilde{e}_2}{dt} \end{bmatrix} + A_C q_t(A_C^T \Pi_e^{-1} \tilde{e}, t) + A_R r(A_R^T \Pi_e^{-1} \tilde{e}, t) + A_L \Pi_j^{-1} \tilde{j}_L + \\ &\qquad A_V j_V + A_I i(\,\cdot\,), \\ 0 &= L(j_L, t) \Pi_j^{-1} \begin{bmatrix} \hat{j}_1 \\ \frac{d\tilde{j}_{L_2}}{dt} \end{bmatrix} + \phi_t(\Pi_j^{-1} \tilde{j}_L, t) - A_L^T \Pi_e^{-1} \tilde{e}, \\ 0 &= A_V^T \Pi_e^{-1} \tilde{e} - v(\,\cdot\,), \\ 0 &= Q_{CRV}^T (A_L \Pi_j^{-1} \begin{bmatrix} \hat{j}_1 \\ \frac{d\tilde{j}_{L_2}}{dt} \end{bmatrix} + \frac{di(\,\cdot\,)}{dt}), \\ 0 &= \bar{Q}_{V-C}^T (A_V^T \Pi_e^{-1} \begin{bmatrix} \hat{e}_1 \\ \frac{d\tilde{e}_2}{dt} \end{bmatrix} - \frac{dv(\,\cdot\,)}{dt}) \end{aligned} \tag{46}$$

is strangeness-free.

Proof. The renaming of the variables gives rise to the same elimination procedure as used in [8] to show that the original problem has differentiation index 2. Thus, the same proof shows the present claim but without the need of differentiating. □

Remark 2 If the original system has size n and there are n_2 equations in (43) then the extended system has size $n+n_2$. Since typically n_2 is much smaller than n, the extended system is only slightly larger than the original system.

For the charge oriented MNA (42) the equations that are responsible for a strangeness index higher than 0 are given by the projected equations in (43) together with the last two equations in (42).

Using the replacements as in (45) and in addition

$$\hat{q} = \frac{dq}{dt}, \quad \hat{\phi} = \frac{d\phi}{dt}, \tag{47}$$

we obtain the following minimally extended system.

$$\begin{aligned}
0 &= A_C\hat{q} + A_R r(A_R^T\Pi_e^{-1}\tilde{e}, t) + A_L\Pi_j^{-1}\tilde{j}_L + A_V j_V + A_I i(\,\cdot\,),\\
0 &= \hat{\phi} - A_L^T\Pi_e^{-1}\tilde{e},\\
0 &= A_V^T\Pi_e^{-1}\tilde{e} - v(\,\cdot\,),\\
0 &= q - q_C(A_C^T\Pi_e^{-1}\tilde{e}, t),\\
0 &= \phi - \phi_L(\Pi_j^{-1}\tilde{j}_L, t),\\
0 &= Q_{CRV}^T(A_L\Pi_j^{-1}\begin{bmatrix} \hat{\tilde{j}}_1 \\ \frac{d\tilde{j}_{L_2}}{dt} \end{bmatrix} + \frac{di(\,\cdot\,)}{dt}),\\
0 &= \bar{Q}_{V-C}^T(A_V^T\Pi_e^{-1}\begin{bmatrix} \hat{e}_1 \\ \frac{d\tilde{e}_2}{dt} \end{bmatrix} - \frac{dv(\,\cdot\,)}{dt}),\\
0 &= \hat{q} - C(A_C^T\Pi_e^{-1}\tilde{e}, t)A_C^T\Pi_e^{-1}\begin{bmatrix} \hat{e}_1 \\ \frac{d\tilde{e}_2}{dt} \end{bmatrix} + q_t(A_C^T\Pi_e^{-1}\tilde{e}, t),\\
0 &= \hat{\phi} - L(\Pi_j^{-1}\tilde{j}_L, t)\Pi_j^{-1}\begin{bmatrix} \hat{\tilde{j}}_1 \\ \frac{d\tilde{j}_{L_2}}{dt} \end{bmatrix} + \phi_t(\Pi_j^{-1}\tilde{j}_L, t).
\end{aligned} \tag{48}$$

Remark 3 Obviously, we can use the last two relations to eliminate the just introduced variables $\hat{q}$ and $\hat{\phi}$ obtaining just the minimally extended system (46) for the conventional MNA. Hence, system (48) is strangeness-free as well. Moreover, from a numerical point of view the reduced problems and minimally extended systems belonging to the conventional and charge oriented MNA are the same or at least equivalent (in the sense that the common part of the numerical solution would be same when using the same stepsizes and ignoring roundoff errors). Concerning efficiency, however, we observe that in the charge oriented MNA the minimally extended strangeness-free system is often significantly larger than the original system.

5 Multi-body systems

A second important class of problems where it is possible to use the structure to derive the reduced derivative array and the minimally extended strangeness-free system are the models for mechanical multi-body systems [32]. The approach that we consider here has been discussed in the context of industrial simulation codes in detail in [1]. For this reason we present this case only very briefly.

The classical first order form of a multi-body system, see, e. g., [7, 29], is

$$\begin{aligned} \dot{p} &= v, \\ M\dot{v} &= f(p,v) - g_p(p)^T\lambda, \\ 0 &= g(p), \end{aligned} \tag{49}$$

where p are the positions, v the velocities, M is the mass matrix, $g(p)$ describes the constraints and λ is the associated Lagrange multiplier. Under the usual assumptions, i. e., that M is positive definite and that the Jacobian $g_p(p)$ has full row rank, this system has differentiation index 3 (or strangeness index 2).

A well-known index reduction technique is given by the Gear-Gupta-Leimkuhler stabilization [9], that couples the time-derivative of the constraint equations via further Lagrange multipliers ν into the dynamics and gives the system of differentiation index 2

$$\begin{aligned} \dot{p} &= v - g_p(p)^T\nu, \\ M\dot{v} &= f(p,v) - g_p(p)^T\lambda, \\ 0 &= g(p), \\ 0 &= g_p(p)v. \end{aligned} \tag{50}$$

It follows that this stabilization also introduces new variables and is therefore an extended system, but it is not strangeness-free. It is clear that we have to perform one more differentiation of the constraint equations to obtain the reduced derivative array as

$$\begin{aligned} \dot{p} &= v, \\ M\dot{v} &= f(p,v) - g_p(p)^T\lambda, \\ 0 &= g(p), \\ 0 &= g_p(p)v, && (51) \\ 0 &= g_{pp}(p)(v,v) + g_p(p)\dot{v}. && (52) \end{aligned}$$

To obtain the minimally extended strangeness-free system we (locally) determine a permutation matrix Π such that for the Jacobian matrix $g_p(p)$ we have

$$g_p(p)\Pi = [G_1 \ G_2],$$

with G_2 being square and nonsingular. We then partition

$$\Pi^T p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}, \quad \Pi^T v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

conformally and replace every occurrence of $\dot{p}_2$ by the new variable w_1 and every occurrence of $\dot{v}_2$ by the new variable w_2. This gives the extended system

$$\begin{aligned} \dot{p}_1 &= v_1, && (53) \\ w_1 &= v_2, && (54) \\ M\Pi \begin{bmatrix} \dot{v}_1 \\ w_2 \end{bmatrix} &= f(p,v) - g_p(p)^T\lambda, && (55) \end{aligned}$$

$$
\begin{aligned}
0 &= g(p), && (56)\\
0 &= g_p(p)v, && (57)\\
0 &= g_{pp}(p)(v,v) + g_p(p)\Pi \begin{bmatrix} \dot{v}_1 \\ w_2 \end{bmatrix}. && (58)
\end{aligned}
$$

The following theorem shows that it is strangeness free.

Theorem 7 *Let M be positive definite and let $g_p(p)$ have full row rank. Then the extended system (53)–(58) is strangeness-free.*

Proof. Since G_2 is square nonsingular we can solve (56) for p_2 in terms of p_1 and (57) for v_2 in terms of p_1 and v_1. Since M is positive definite and $g_p(p)$ has full row rank, it follows that also $H(p) = g_p(p)M^{-1}g_p(p)^T$ is positive definite and hence we can use (58) to obtain

$$\lambda = H(p)^{-1}(g_{pp}(p)(v,v) + g_p(p)M^{-1}f(p,v)).$$

Finally, the positive definiteness of M implies that we can solve for $\dot{v}_1$ and w_2 and it remains an ordinary differential equation in the unknowns p_1 and v_1. Thus, the system has strangeness index 0. □

Remark 4 If the original system (49) has n_p dynamical equations and n_c constraints, then the minimally extended strangeness-free system consists of $n_p + 3n_c$ equations in the same number of unknowns. Since typically the number of constraints is much smaller than the number of dynamical equations, this approach is feasible from a complexity point of view, in particular, in view of the fact that the resulting system is strangeness-free and can be treated by all integrators for systems of differentiation index 1. We will demonstrate this in Section 6.

6 Numerical examples

To demonstrate the gain in efficiency that can be obtained when working with the minimally extended strangeness-free system, we discuss two examples. All computations were performed on a Sun Ultra-1 workstation under Fortran 77, using the GNU Fortran compiler.

First consider the system

$$
\begin{aligned}
J + G(u_1 - u_2) &= 0,\\
C_1\dot{u}_2 - G(u_1 - u_2) + C_2(\dot{u}_2 - \dot{u}_3) &= 0,\\
J_V - C_2(\dot{u}_2 - \dot{u}_3) &= 0,\\
u_1 - V(t) &= 0,\\
u_3 - Au_2 &= 0, \qquad (59)
\end{aligned}
$$

taken from [12], modeling the so-called Miller integrator circuit. For parameter values $C_1 = C_2 = 1$, $A = 1$, and $G = 1$, this system is known to have differentiation index 2, hence strangeness index 1. Obviously, the higher index is caused by the coupling between u_3 and $\dot{u}_3$. The minimally extended strangeness-free system is obtained by adding the differentiated relation $\dot{u}_3 - A\dot{u}_2 = 0$ and replacing $\dot{u}_3$ say by $\hat{u}_3$. This leads to the system

$$\begin{aligned} J + G(u_1 - u_2) &= 0, \\ C_1\dot{u}_2 - G(u_1 - u_2) + C_2(\dot{u}_2 - \hat{u}_3) &= 0, \\ J_V - C_2(\dot{u}_2 - \dot{u}_3) &= 0, \\ u_1 - V(t) &= 0, \\ u_3 - Au_2 &= 0, \\ \hat{u}_3 - A\dot{u}_2 &= 0. \end{aligned} \tag{60}$$

From this we can get a reduced strangeness-free problem by eliminating $\hat{u}_3$, hence

$$\begin{aligned} J + G(u_1 - u_2) &= 0, \\ C_1\dot{u}_2 - G(u_1 - u_2) + C_2(\dot{u}_2 - A\dot{u}_2) &= 0, \\ J_V - C_2(\dot{u}_2 - A\dot{u}_2) &= 0, \\ u_1 - V(t) &= 0, \\ u_3 - Au_2 &= 0. \end{aligned} \tag{61}$$

In Table 1 we present the CPU-times needed for solving these systems with the code GELDA [22] with tolerance 10^{-5} and 10^{-9}.

Tolerance	standard	str.-free reduced	str.-free min. ext.
10^{-5}	1.61	0.29	0.42
10^{-9}	7.03	1.27	1.91

Table 1: Runtime in seconds for Miller circuit

We see that much can be gained by using the strangeness-free reduced form, but the computational effort that is needed for the strangeness-free minimally extended system is not significantly higher. However, it is in general easier to get this form than the fully reduced system.

The second example is a multi-body system describing the movement of a mass point restricted to a parabola under gravity. The equations taken from [31] have the form

$$\begin{aligned} \dot{p}_1 &= v_1,\ \dot{p}_2 = v_2,\ \dot{p}_3 = v_3, \\ \dot{v}_1 &= 2\lambda p_1, \\ \dot{v}_2 &= 2\lambda p_2, \\ \dot{v}_3 &= -\lambda - 1, \\ 0 &= p_1^2 + p_2^2 - p_3. \end{aligned} \tag{62}$$

Here the coupling between p_3 and $\dot{p}_3$ causes a higher index. Differentiating the constraint once and eliminating the differentiated variables with the help of the other equations yields

$$0 = 2p_1v_1 + 2p_2v_2 - v_3.$$

Now the coupling between v_3 and $\dot{v}_3$ causes a higher index. Differentiating once more and eliminating gives

$$0 = 2v_1^2 + 4\lambda p_1^2 + 2v_2^2 + 4\lambda p_2^2 + \lambda + 1.$$

According to Section 5 a minimally extended strangeness-free system is obtained by putting the above equations together and replacing $\dot{p}_3$ and $\dot{v}_3$ say by $\hat{p}_3$ and $\hat{v}_3$. The system then reads

$$\begin{aligned}
\dot{p}_1 &= v_1,\ \dot{p}_2 = v_2,\ \hat{p}_3 = v_3,\\
\dot{v}_1 &= 2\lambda p_1,\\
\dot{v}_2 &= 2\lambda p_2,\\
\hat{v}_3 &= -\lambda - 1,\\
0 &= p_1^2 + p_2^2 - p_3,\\
0 &= 2p_1v_1 + 2p_2v_2 - v_3,\\
0 &= 2v_1^2 + 4\lambda p_1^2 + 2v_2^2 + 4\lambda p_2^2 + \lambda + 1.
\end{aligned} \tag{63}$$

A reduced strangeness-free system is achieved by simply omitting the equations that involve the variables $\hat{p}_3$ and $\hat{v}_3$. Hence, we have

$$\begin{aligned}
\dot{p}_1 &= v_1,\ \dot{p}_2 = v_2,\\
\dot{v}_1 &= 2\lambda p_1,\\
\dot{v}_2 &= 2\lambda p_2,\\
0 &= p_1^2 + p_2^2 - p_3,\\
0 &= 2p_1v_1 + 2p_2v_2 - v_3,\\
0 &= 2v_1^2 + 4\lambda p_1^2 + 2v_2^2 + 4\lambda p_2^2 + \lambda + 1.
\end{aligned} \tag{64}$$

In Table 2 we present a comparison of the results obtained when solving these systems by the code GENDA [23] and with the specialized multi-body code ODASSL of Führer, see [7], applied to the original system.

Looking at Table 2 we see that much can be gained in the computational effort when using the analytically produced strangeness-free version and the minimally extended system is only a factor 2 more expensive. However, the special solver for multi-body systems is still more efficient, the main reason being that much fewer factorizations are needed in ODASSL than in the other solvers. This is due to the fact that the factorization is kept fixed for several steps, while in GENDA a new factorization is determined at every step. Even though it is preferable to use a special code like ODASSLXS if the structure is known, we have demonstrated that the general purpose code GENDA can be made almost as efficient as the well established code ODASSL.

	ODASSL	GENDA standard	GENDA str.-fr. reduced	GENDA str.-fr. min. ext.
steps	4536	5136	5201	5326
calls DRES	13097	20352	10402	10652
factorizations	1351	5136	5201	5326
fails error test	2	0	0	0
fails conv. test	0	0	0	0
runtime (sec)	0.07	2.97	0.52	0.93

Table 2: Comparison of different formulations computed with GENDA and ODASSL. Tolerance RTOL=ATOL=10^{-9}

7 Conclusion

We have discussed index reduction methods for large scale differential-algebraic systems, where the structure of the problem can be used to identify the equations of the systems that are responsible for higher index. In order to avoid expensive subspace and rank computations, new variables are introduced that extend the system size in a minimal way leading to a system that is strangeness-free (or if defined of differentiation index 1).

We have demonstrated this new approach for circuit simulation and for multibody systems, and we have given numerical examples that show the effectiveness of the approach.

8 Acknowledgment

We thank Diana Estevez-Schwarz and Caren Tischendorf for many helpful discussions concerning the topological characterization of the high index equations in circuit simulation and Ingo Seufer for running the test cases in Section 6. We also thank an anonymous referee for pointing out some inaccuracies in a previous version of the paper.

References

[1] M. Arnold, V. Mehrmann, and A. Steinbrecher. Index reduction in industrial multibody system simulation. Technical Report DLR IB 532-01-01, DLR German Aerospace Center, Institute of Aeroelasticity, Vehicle Systems Group, P.O. Box 1116, D-82230 Wessling, FRG, 2001.

[2] A. Barrlund and B. Kågström. Analytical and numerical solutions to higher index linear variable coefficient DAE systems. *J. Comput. Appl. Math.*, 31:305–330, 1990.

[3] B. Benhammouda. *Numerical Solution of Large Differential-Algebraic Systems on massively Parallel Computers.* PhD thesis, Fakultät für Mathematik, TU Chemnitz, D-09107 Chemnitz, 1998. Dissertationsschrift.

[4] K. E. Brenan, S. L. Campbell, and L. R. Petzold. *Numerical Solution of Initial-Value Problems in Differential Algebraic Equations*, volume 14 of *Classics in Applied Mathematics*. SIAM, Philadelphia, PA, second edition, 1996.

[5] S. L. Campbell. Comment on controlling generalized state-space (descriptor) systems. *Internat. J. Control*, 46:2229–2230, 1987.

[6] S. L. Campbell and C. W. Gear. The index of general nonlinear DAEs. *Numer. Math.*, 72(2):173–196, 1995.

[7] E. Eich-Soellner and C. Führer. *Numerical Methods in Multibody Systems*. B. G. Teubner Stuttgart, 1998.

[8] D. Estévez-Schwarz and C. Tischendorf. Structural analysis for electrical circuits and consequences for MNA. *Int. J. Circ. Theor. Appl.*, 28:131–162, 2000.

[9] C. W. Gear, B. Leimkuhler, and G. K. Gupta. Automatic integration of Euler-Lagrange equations with constraints. *J. Comput. Appl. Math.*, 12/13:77–90, 1985.

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

[11] M. Günther and U. Feldmann. CAD-based electric-circuit modeling in industry I. Mathematical structure and index of network equations. *Surv. Math. Ind.*, 8:97–129, 1999.

[12] M. Günther and U. Feldmann. CAD-based electric-circuit modeling in industry II. Impact of circuit configurations and parameters. *Surv. Math. Ind.*, 8:131–157, 1999.

[13] E. Hairer, C. Lubich, and M. Roche. *The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods*. Lecture Notes in Mathematics No. 1409. Springer-Verlag, Berlin, 1989.

[14] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Springer Verlag, Berlin, second edition, 1996.

[15] P. Kunkel and V. Mehrmann. Canonical forms for linear differential-algebraic equations with variable coefficients. *J. Comput. Appl. Math.*, 56:225–259, 1994.

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

[17] P. Kunkel and V. Mehrmann. A new class of discretization methods for the solution of linear differential algebraic equations with variable coefficients. *SIAM J. Numer. Anal.*, 33(5):1941–1961, 1996.

[18] P. Kunkel and V. Mehrmann. The linear quadratic control problem for linear descriptor systems with variable coefficients. *Math. Control, Signals, Sys.*, 10:247–264, 1997.

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

[20] P. Kunkel and V. Mehrmann. Analysis of over- and underdetermined nonlinear differential-algebraic systems with application to nonlinear control problems. *Math. Control, Signals, Sys.*, 14:233–256, 2001.

[21] P. Kunkel, V. Mehrmann, and W. Rath. Analysis and numerical solution of control problems in descriptor form. *Math. Control Signals, Sys.*, 14:29–61, 2001.

[22] P. Kunkel, V. Mehrmann, W. Rath, and J. Weickert. Gelda: A software package for the solution of general linear differential algebraic equations. *SIAM J. Sci. Comput.*, 18:115 – 138, 1997.

[23] P. Kunkel, V. Mehrmann, and I. Seufer. Genda: A software package for the numerical solution of general nonlinear differential-algebraic equations. preprint 730, Institut für Mathematik, TU Berlin, Str. des 17. Juni 136, D-10623 Berlin, FRG, 2002.

[24] R. Lamour, R. März, and R.M.M. Mattheij. On the stability behaviour of systems obtained by index reduction. *J. Comput. Appl. Math.*, 56:305–319, 1994.

[25] S. Mattsson and G. Söderlind. Index reduction in differential-algebraic equations using dummy derivatives. *SIAM J. Sci. Statist. Comput.*, 14:677–692, 1993.

[26] J. Ortega and W. Rheinboldt. *Iterative Solutions of Nonlinear Equations in Several Variables.* Academic Press, New York, 1970.

[27] C. C. Pantelides. The consistent initialization of differential-algebraic systems. *SIAM J. Sci. Statist. Comput.*, 9:213–231, 1988.

[28] P. J. Rabier and W. C. Rheinbold. A geometric treatment of implicit differential-algebraic equations. *J. Differential Equations*, 109:110–146, 1994.

[29] P. J. Rabier and W. C. Rheinboldt. *Nonholonomic Motion of Rigid Mechanical Systems from a DAE Viewpoint.* SIAM, Philadelphia, PA 19104-2688, USA, 2000.

[30] G. Reißig, W. S. Martinson, and P. I. Barton. Differential-algebraic equations of index 1 may have an arbitrarily high structural index. To appear in SIAM J. Sci. Comput., 1999.

[31] W. C. Rheinboldt. *Numerical Analysis of Parametrized Nonlinear Equations*. John Wiley & Sons, New York, 1985.

[32] W. Schiehlen. *Multibody Systems Handbook*. Springer-Verlag, 1990.

[33] C. Tischendorf. Feasibility and stability behavior of the BDF applied to index-2 differential algebraic equations. *Z. Angew. Math. Mech.*, 12:927–946, 1995.

[34] C. Tischendorf. Topological index calculation of differential-algebraic equations in circuit simulation. *Surveys on Mathematics in Industry*, 8:187–199, 1999.