Formulating differential algebraic equations properly

Inmaculada Higueras*
Universidad Pública de Navarra
Departamento de Matemática e Informática
ES-31006 Pamplona, Spain

Roswitha März
Humboldt-University Berlin
Institute of Mathematics
Unter den Linden 6
D-10099 Berlin, Germany

Abstract: When modeling complex processes by means of DAEs, one should carefully investigate the leading term. In this paper we study properly formulated problems, i.e. problems of the form $f((Dx)', x, t) = 0$ with some additional condition on $\operatorname{Im} D(t)$. From the numerical point of view, working with properly formulated problems in some sense means that the ODE method is used to discretize only the differential components. We provide stability inequalities and prove convergence in compact integration intervals for Runge-Kutta methods and BDFs. The case of a time invariant $\operatorname{Im} D(t)$ is essentially favourable for the qualitative behaviour of the approximation on infinite intervals. This study is extended to the nonlinear version $f((d(x,t))', x, t) = 0$.

Keywords: Differential algebraic equation, Runge-Kutta methods, backward differentiation formulas

AMS subject classifications: 34A09, 65L80, 65L06

1 Introduction

Although extensive literature on the numerical treatment of DAEs of index 1 has been available for years ([8], [9], [3]), so far, the problem of the correct reflection of qualitative solution properties by numerical approximations have been solved only insufficiently.
For DAEs, it would be desirable to obtain results on the dynamic system behaviour, which are analogous or similar to those available for autonomous regular ODEs (e.g. [14]). If we assume linear homogeneous DAEs with constant coefficients

$$Ax'(t) + Bx(t) = 0$$

*This paper has been done during the stay at the Institute of Mathematics at the Humboldt-University Berlin, supported by a grant from the "Secretaría de Estado de Universidades, Investigación y Desarrollo"

as test equations, the linear stability theory of regular ODEs can be formally applied [8]. However, the problem with DAEs is that time-varying coefficients have a much stronger impact than in case of regular ODEs. In comparison, the class of the well understood linear constant coefficient DAEs is, within the framework of DAEs, not nearly as important as the regular constant coefficient ODEs within the framework of regular ODE-theory.
The simple academic example [10] makes this very obvious.

Example 1.1 *The DAE*

$$\begin{pmatrix} \delta-1 & \delta t \\ 0 & 0 \end{pmatrix} x'(t) + \sigma \begin{pmatrix} \delta-1 & \delta t \\ \delta-1 & \delta t-1 \end{pmatrix} x(t) = 0 \tag{1.1}$$

has index 1 on $\mathbb{R}$ for arbitrary parameters $\sigma \neq 0, \delta \neq 1$. Its solution is

$$x_1(t) = \frac{1-\delta t}{\delta-1} x_2(t), \quad x_2(t) = e^{(\delta-\sigma)t} x_{2,0}\,.$$

For $\delta < \sigma$, the solution is exponentially stable, and one would expect the B-stable implicit Euler method to reflect this behaviour well independently of the chosen stepsize. However, the implicit Euler method applied to (1.1) leads to

$$x_{1,n+1} = \frac{1-\delta t_{n+1}}{\delta-1} x_{2,n+1}, \quad x_{2,n+1} = \frac{1+\delta h}{1+\sigma h} x_{2,n}\,.$$

For $\sigma = 0$, i.e. in the constant coefficient case, all things are fine. But obviously, if $\delta \neq 0$, the condition $\delta < \sigma$ does not imply $|1+\delta h| < |1+\sigma h|$ without stepsize restrictions.
Observe that close to $\sigma h \approx -1$ the numerical approximation explode while with the true solution nothing happens.

Considering linear variable coefficient DAEs

$$A(t)x'(t) + B(t)x(t) = q(t) \tag{1.2}$$

and nonlinear equations

$$A(t)x'(t) + b(x(t), t) = 0\,,$$

one is confronted with the question about the precise meaning of the term $A(t)x'(t)$ and therefore how it should be discretized. In particular, we would like to know the right function class where the solution should belong to. Clearly, for a constant coefficient $A(t) \equiv A$, we may interpret

$$Ax'(t) := (Ax(t))' = A(A^+Ax(t))' = AA^+(Ax(t))'\,,$$

where A^+ denotes the Moore-Penrose inverse. Note that $\operatorname{Im} AA^+ = \operatorname{Im} A$, $\operatorname{Ker} A^+A = \operatorname{Ker} A$.

However, the situation is more complicated if $A(t)$ varies with time. In [8], a projector function $P(t)$ along the nullspace $\operatorname{Ker} A(t)$ (i.e. $P(t)^2 = P(t)$, $\operatorname{Ker} P(t) = \operatorname{Ker} A(t)$), e.g. $P(t) = A(t)^+A(t)$, is introduced and the meaning of $A(t)x'(t)$ is given by

$$A(t)x'(t) = A(t)P(t)x'(t) = A(t)\left\{P(t)x(t))' - P'(t)x(t)\right\}.$$

Hence (1.2) is reformulated as

$$A(t)(P(t)x(t))' + (B(t) - A(t)P'(t))x(t) = q(t)\,. \tag{1.3}$$

Note that, if the nullspace $\operatorname{Ker} A(t)$ is constant, we may choose a constant $P(t) \equiv P$, such that

$$A(t)x'(t) = A(t)Px'(t) = A(t)(Px(t))',$$

and the derivative $P'(t)$ in (1.3) disappears. In this case, the numerical integration methods are known to work well as expected from the regular ODE view point [8]. Note that in our above example (1.1) the subspace $\operatorname{Ker} A(t)$ varies with t.

A further interpretation of the leading term in (1.2) is (e.g. [7])

$$A(t)x'(t) := (A(t)x(t))' - A'(t)x(t)$$

what leads to the reformulation of (1.2) as

$$(A(t)x(t))' + (B(t) - A'(t))x(t) = q(t)\,. \tag{1.4}$$

Example 1.2 *Reformulating (1.1) into (1.4)*

$$\left\{\begin{pmatrix} \delta-1 & \delta t \\ 0 & 0 \end{pmatrix} x(t)\right\}' + \left\{\sigma \begin{pmatrix} \delta-1 & \delta t \\ \delta-1 & \delta t-1 \end{pmatrix} - \begin{pmatrix} 0 & \delta \\ 0 & 0 \end{pmatrix}\right\} x(t) = 0\,,$$

the implicit Euler method leads to

$$x_{1,n+1} = \frac{1-\delta t_{n+1}}{\delta-1} x_{2,n+1}\,, \quad x_{2,n+1} = \frac{1}{1-(\delta-\sigma)h} x_{2,n}\,.$$

Now we are lucky. For $\delta - \sigma < 0$, without any stepsize restriction, the numerical approximation x_{n+1} tends to zero as the true solution $x(t_{n+1})$ does.

By means of a projector function $R(t)$ onto $\operatorname{Im} A(t)$ (i.e. $R(t)^2 = R(t)$, $\operatorname{Im} A(t) = \operatorname{Im} R(t)$), say $R(t) = A(t)A(t)^+$, one may also use

$$A(t)x'(t) = R(t)A(t)x'(t) = R(t)\left\{(A(t)x(t))' - A'(t)x(t)\right\}$$

yielding

$$R(t)(A(t)x(t))' + (B(t) - R(t)A'(t))x(t) = q(t)\,. \tag{1.6}$$

In [7], a particular well behaviour of numerical approximations for (1.4) is reported for time invariant subspaces $\operatorname{Im} A(t)$. In this case we may use a constant $R(t) \equiv R$, so that the formulations (1.6) and (1.4) are numerically equivalent, i.e. , they give the same numerical solution. Recall that this is the case for the above example.
Observe that the leading terms in both reformulations (1.3) and (1.6) are given in terms of two matrix functions, namely $A(t)$, $P(t)$ and $R(t)$, $A(t)$ respectively.
At this place, it should be mentioned that the adjoint equation to (1.2) resp. (1.3) has the form ([2])

$$(A^*(t)y(t))' - B^*(t)y(t) = 0\,,$$

or more precisely

$$P^*(t)(A^*(t)y(t))' - (B^*(t) - P^*(t)A^{*'}(t))y(t) = 0\,.$$

Again, the leading term is given by two matrix functions.

It comes out that to formulate the leading term of a DAE properly we should consider two matrix functions. None of them has to be necessarily a projector function but both have to be well matched together in some sense. This concept is used in [1], [11], where linear equations

$$A(t)(D(t)x(t))' + B(t)x(t) = q(t) \tag{1.7}$$

and their adjoint equations are analyzed.
It is worth mentioning that in very early papers concerning the numerical integration of DAEs (e.g. [12]) leading terms as in (1.7) were considered. In particular, different versions of linear multistep methods applied to nonlinear DAEs of index 1 are shown to be stable and convergent.

In the present paper we take off this approach. In Section 2 we formulate and analyze the DAE

$$A(x(t),t)(D(t)x(t))' + b(x(t),t) = 0 \tag{1.8}$$

by means of an appropriate generalization of the decoupling technique proposed in [8]. In particular, while in [8], due to different projectors, a variety of inherent regular ODEs had to be taken into account, now the inherent regular ODE is uniquely determined by the problem data.
Runge-Kutta (RK) methods and Backward Differentiation Formulas (BDFs) are applied to general index-1 DAEs in Section 3. We provide stability inequalities and show convergence in the case of a compact integration interval.
In Section 4 we deal with the question whether decoupling and discretization commute. A positive answer has nice consequences for the qualitative behaviour of the approximation on infinite intervals. This is why we call those DAEs numerically

well formulated.
Finally, in Section 5 we give a further generalization when $(D(t)x(t))'$ in the leading term of (1.8) is replaced by a possible nonlinear version $(d(x(t),t))'$. Some conclusions and forthcoming work are pointed out in Section 6.

2 DAEs with properly formulated leading term and the decoupling in the index-1-case

Consider equations of the form

$$A(x(t),t)(Dx)'(t) + b(x(t),t) = 0 \tag{2.1}$$

where $A : \mathcal{D}_0 \times \mathcal{I}_0 \subseteq I\!R^m \times I\!R \longrightarrow L(I\!R^m), b : \mathcal{D}_0 \times \mathcal{I}_0 \longrightarrow I\!R^m$ are continuous with continuous partial derivatives A'_x, b'_x, and $D(t) \in L(I\!R^m)$ depends continuously on $t \in \mathcal{I}_0$.
For brevity, instead of (2.1) we sometimes write

$$f((Dx)'(t), x(t), t) = 0 \tag{2.2}$$

with $f(y,x,t) := A(x,t)y + b(x,t)$.
A function $x(.) : \mathcal{I}_x \longrightarrow I\!R^m$ is said to be a solution of (2.1) in the interval $\mathcal{I}_x \subseteq \mathcal{I}_0$, if it belongs to the function space

$$C^1_D(\mathcal{I}_x, I\!R^m) := \{x(.) \in C(\mathcal{I}_x, I\!R^m) : (Dx)(.) \in C^1(\mathcal{I}_x, I\!R^m)\}$$

and (2.1) is satisfied for all $t \in \mathcal{I}_x$.
Obviously, with

$$\mathcal{M}_0(t) := \{x \in \mathcal{D}_0 : b(x,t) \in \operatorname{Im} A(x,t)\}, \qquad t \in \mathcal{I}_0,$$

we have for all solutions that

$$x(t) \in \mathcal{M}_0(t), \quad t \in \mathcal{I}_x\,.$$

If $A(x,t)$ and $D(t)$ remain everywhere nonsingular, then (2.1) simply represents a regular ODE with respect to $D(t)x(t)$, and $\mathcal{M}_0(t) = \mathcal{D}_0$ ([5]). We are interested in the more complicated case of DAEs characterized by everywhere singular matrices $A(x,t)$ and $D(t)$.

Definition 2.1 *The DAE (2.1) has a properly formulated leading term, if*

$$\operatorname{Ker} A(x,t) \oplus \operatorname{Im} D(t) = I\!R^m \text{ for all } x \in \mathcal{D}_0, t \in \mathcal{I}_0, \tag{2.3}$$

and if there is a $R \in C^1(\mathcal{I}_0, L(I\!R^m))$ *such that*
$R(t)^2 = R(t)$, $\operatorname{Im} R(t) = \operatorname{Im} D(t)$, $\operatorname{Ker} R(t) = \operatorname{Ker} A(x,t)$, *for all* $t \in \mathcal{I}_0, x \in \mathcal{D}_0$.
The matrix functions $A(x,t)$ *and* $D(t)$ *are said to be well matched.*

Observe that the conditions on $R(t)$ imply that $A(x,t)R(t) = A(x,t)$, $f(y,x,t) = f(R(t)y,x,t)$ and $R(t)D(t) = D(t)$.

Remark 2.1 *In particular, if the leading term of (2.1) is properly formulated,* $\operatorname{Ker} A(x,t)$ *is independent of* x*, and both* $\operatorname{Im} D(t)$ *and* $\operatorname{Ker} A(x,t)$ *have constant dimension, say* r *and* $m-r$*. Furthermore, both subspaces are spanned by* C^1 *functions.*

Remark 2.2 *Condition (2.3) is equivalent to the three relations*

$$\operatorname{Im} A(x,t)D(t) = \operatorname{Im} A(x,t)\,,$$
$$\operatorname{Ker} A(x,t)D(t) = \operatorname{Ker} D(t)\,,$$
$$\operatorname{Ker} A(x,t) \cap \operatorname{Im} D(t) = \{0\}, \quad x \in \mathcal{D}_0,\ t \in \mathcal{I}_0\,.$$

Remark 2.3 *The (reformulated) DAEs (1.3) and (1.6) have properly formulated leading terms. In general, in a properly formulated DAE, the matrices defining the leading term are well matched together, and the leading term shows precisely all involved derivatives.*

For the rest of the paper we introduce the following definitions and notations:

$N_0(t) := \operatorname{Ker} D(t)$,

$Q_0(t)$ is a projector onto $N_0(t), P_0(t) := I - Q_0(t)$,

$B(y,x,t) := b'_x(x,t) + (A(x,t)y)'_x$,

$S_0(y,x,t) := \{z \in I\!R^m : B(y,x,t)z \in \operatorname{Im} A(x,t)\}$,

$A_1(y,x,t) := A(x,t)D(t) + B(y,x,t)Q_0(t)$.

Observe that by construction, it holds that $B(y,x,t) = B(R(t)y,x,t)$.
If the DAE (2.1) has a properly formulated leading term, then for each $x_0 \in \mathcal{M}_0(t)$ there is a unique $y_0 \in \operatorname{Im} D(t)$ such that $A(x_0,t)y_0 + b(x_0,t) = 0$. Then $S_0(y_0,x_0,t) = T_{x_0}\mathcal{M}_0(t)$ holds true.
Further, we denote by $D(t)^-$ the reflexive generalized inverse of $D(t)$ that has the additional properties

$$D(t)D(t)^- = R(t)\,, \qquad D(t)^- D(t) = P_0(t)\,.$$

Obviously, $D(t)^-$ depends on how $P_0(t)$ is chosen. Observe that $D(t)P_0(t) = D(t)$ and hence $D(t)Q_0(t) = 0$. Observe too that $P_0(t)D(t)^- = D(t)^-$, and hence $Q_0(t)D(t)^- = 0$.

In the following, throughout this paper, we assume the DAE (2.1) to be properly formulated in the sense that the matrices defining its leading term are well matched. It should be stressed that this has nothing to do with the mathematical notion of well-posedness of a problem. It only says that the derivatives involved in fact are figured out in a proper way.

We begin with the definition of DAE tractable with index-1.

Definition 2.2 *DAE (2.1) is said to be tractable with index 1, if*

$$N_0(t) \cap S_0(y,x,t) = \{0\}, \quad x \in \mathcal{D}_0\,,\; t \in \mathcal{I}_0\,,\; y \in \mathbb{R}^m\,. \tag{2.4}$$

Remark 2.4 *It would be enough to impose $N_0(t) \cap T_{x_0}\mathcal{M}_0(t) = \{0\}$ for all $x \in \mathcal{M}_0(t)$, that, due to the continuity of $A_1(y,x,t)$ with respect to its arguments, implies immediately $N_0(t) \cap S_0(y,x,t) = \{0\}$ on a neighbourhood. One can interpret $\mathcal{D}_0$ to represent this neighbourhood.*

Remark 2.5 *Using [8, Th. 13, p. 198], the relation (2.4) is equivalent to any of the following conditions:*

1. *$A_1(y,x,t)$ is nonsingular for all $x \in \mathcal{D}_0$, $t \in \mathcal{I}_0$, $y \in \mathbb{R}^m$.*
2. *$N_0(t) \oplus S_0(y,x,t) = \mathbb{R}^m$ for all $x \in \mathcal{D}_0$, $t \in \mathcal{I}_0$, $y \in \mathbb{R}^m$.*

The projector onto $S_0(y,x,t)$ along $N_0(t)$ is called the canonical projector in the index 1 case and will be denoted by $P_s(y,x,t)$. An useful representation ([8]) is

$$P_s = I - Q_0 A_1^{-1} B\,.$$

For any vector x, we can write

$$x = P_0(t)x + Q_0(t)x = D(t)^- D(t)x + Q_0(t)x\,.$$

In this way (2.2) can be expressed as

$$f(R(t)(D(t)x(t))', D(t)^- D(t)x + Q_0(t)x, t) = 0\,. \tag{2.5}$$

Denoting $w(t) = D(t)^-(D(t)x(t))' + Q_0(t)x(t)$, as $D(t)w(t) = R(t)(D(t)x(t))'$, $Q_0(t)w(t) = Q_0(t)x(t)$, we rewrite (2.5) as

$$f\left(D(t)w(t), D(t)^- D(t)x(t) + Q_0(t)w(t), t\right) = 0\,.$$

In the following lemma we study the equation

$$f(D(t)w, D(t)^- u + Q_0(t)w, t) = 0\,.$$

Lemma 2.1 *Given $t_0 \in \mathcal{I}_0$, $x_0 \in \mathcal{M}_0(t_0)$, $y_0 \in \operatorname{Im} D(t_0)$ such that*

$$N(t_0) \cap S(y_0, x_0, t_0) = \{0\}\,, \qquad f(y_0, x_0, t_0) = 0\,,$$

we denote

$$u_0 := D(t_0)x_0 \qquad w_0 := D(t_0)^- y_0 + Q_0(t_0)x_0\,,$$

and define

$$F(w,u,t) := f(D(t)w, D(t)^- u + Q_0(t)w, t)$$

for $(w,u,t) \in \mathcal{N}_0 \subset I\!R^m \times I\!R^m \times I\!R$, *where* $\mathcal{N}_0$ *is a neighbourhood of* (w_0,u_0,t_0). *Then, there is a continuous function*

$$w : B(u_0,\rho) \times \mathcal{I} \longrightarrow I\!R^m, \quad w(u_0,t_0) = w_0$$

that satisfies

$$F(w(u,t),u,t) = 0, \quad u \in B(u_0,\rho), \quad t \in \mathcal{I}\,.$$

It holds that $w(u,t) \equiv w(R(t)u,t)$. *Furthermore* w *has a continuous partial derivative* w'_u *satisfying*

$$\begin{aligned} w'_u(u,t) &= -(A_1^{-1}B)(D(t)w(u,t),\ D(t)^-u + Q_0(t)w(u,t),\ t)D(t)^-, \\ w'_u(u_0,t_0) &= -(A_1^{-1}B)(y_0,x_0,t_0)D(t_0)^-\ . \end{aligned}$$

Proof. Since $F(w_0,u_0,t_0) = 0$, as

$$F'_w(w_0,u_0,t_0) = A(x_0,t_0)D(t_0) + B(y_0,x_0,t_0)Q_0(t_0) = A_1(y_0,x_0,t_0)\,,$$

the assertion results from the Implicit Function Theorem.

□

By Lemma 2.1, the relations

$$f(D(t)w, D(t)^-u + Q_0(t)w, t) = 0 \text{ and } w = w(u,t)$$

are locally equivalent around points $x_0 \in \mathcal{M}_0(t_0)$. Given a solution $x(.) \in C^1_D(\mathcal{I}, I\!R^m)$ of the index-1 tractable DAE (2.1), we may apply Lemma 2.1 at each $x(t) \in \mathcal{M}_0(t), t \in \mathcal{I}$. By uniqueness and continuity arguments we find the function w to be given around $\{(u(t),t) : t \in \mathcal{I}\}$, where

$$u(t) := D(t)x(t), \quad w(t) := D(t)^-(D(t)x(t))' + Q_0(t)x(t).$$

By this way, the following representation of the solution results:

$$x(t) = D^-(t)u(t) + Q_0(t)w(u(t),t), \tag{2.6}$$

where $u(.) \in C^1$ satisfies the equation

$$R(t)u'(t) = D(t)w(u(t),t). \tag{2.7}$$

Since $R(.) \in C^1$ we may rewrite (2.7) into

$$u'(t) - R'(t)u(t) = D(t)w(u(t),t). \tag{2.8}$$

The solution representation (2.6), (2.8) gives a nice insight in the DAE structure. We will refer to (2.8) as the inherent regular ODE. Recall that the flow is governed mainly by the inherent regular ODE.

Obviously, in the context of Lemma 2.1, we may consider the regular ODE (2.8) without assuming the existence of a DAE solution.

Theorem 2.2 *Let all assumptions and the function* $w : B(u_0, \rho) \times \mathcal{I} \longrightarrow \mathbb{R}^m$ *of Lemma 2.1 be given.*

(i) Then, the time varying subspace $\operatorname{Im} D(t) = \operatorname{Im} R(t), t \in \mathcal{I}$, *is an invariant subspace of the inherent regular ODE (2.8) of the DAE (2.1).*
That is, if a solution starts in $u(t_0) \in \operatorname{Im} D(t_0)$ *for some* $t_0 \in \mathcal{I}$, *it holds that* $u(t) \in \operatorname{Im} D(t)$ *for all* t, *where the solution exists.*

(ii) If $\operatorname{Im} D(t)$ *does not vary with* t, *the inherent regular ODE simplifies on the invariant subspace* $\operatorname{Im} D(t_0)$ *to*

$$u'(t) = D(t)w(u(t), t), \quad u(t_0) \in \operatorname{Im} D(t_0).$$

(iii) The inherent regular ODE (2.8) is uniquely determined by the data of the DAE. In particular, (2.8) is independent of the choice of the projector $P_0(t)$.

Proof.

(i) Inserting any solution $u(\cdot) \in C^1$ in (2.8) we multiply the resulting identity by $(I - R(t))$, what leads to

$$(I - R(t))u'(t) - (I - R(t))R'(t)u(t) = 0,$$

or, with $v(t) := (I - R(t))u(t)$,

$$\begin{aligned} v'(t) &= -R'(t)u(t) + (I - R(t))R'(t)u(t) \\ &= -R'(t)u(t) + R'(t)R(t)u(t) \\ &= -R'(t)v(t). \end{aligned}$$

Hence, $v(t)$ vanishes identically if $v(t_0) = 0$.

(ii) Denoting by $\bar{R} \in L(\mathbb{R}^m)$ a constant projector onto the constant subspace $\operatorname{Im} D(t)$ we have $\bar{R} = R(t)\bar{R}$, $R(t) = \bar{R}R(t)$ and

$$R'(t)R(t) = R(t)'\bar{R}R(t) = (R(t)\bar{R})'R(t) = \bar{R}'R(t) = 0.$$

Therefore, due to $u(t) = R(t)u(t)$, the term $R'(t)u(t)$ in (2.8) disappears.

(iii) We consider two projectors P_0, $\check{P}_0$ along N_0, and the corresponding pseudoinverses $D(t)^-$, $\check{D}(t)^-$. We apply Lemma 2.1 using $P_0, D(t)^-$ and $\check{P}_0(t), \check{D}(t)^-$

respectively.
The relation $w = w(u,t)$ is locally equivalent with

$$\begin{aligned} 0 &= f(D(t)w, D(t)^- u + Q_0(t)w, t) \\ &= f(D(t)w, \check{P}_0(t)\{D(t)^- u + Q_0(t)w\} + \check{Q}_0(t)\{D(t)^- u + Q_0(t)w\}, t). \end{aligned}$$

Since

$$\check{P}_0(t)\{D(t)^- u + Q_0(t)w\} = \check{P}_0(t)D(t)^- u = \check{D}(t)^- u\,,$$

this means

$$\begin{aligned} w = w(u,t) \quad &\Leftrightarrow \quad 0 = f(D(t)w, \check{D}(t)^- u + \check{Q}_0(t)\{D(t)^- u + Q_0(t)w\}, t) \\ &\Leftrightarrow \quad \breve{w} := \check{P}_0(t)w + \check{Q}_0(t)\{D(t)^- u + Q_0(t)w\} = \breve{w}(u,t)\,. \end{aligned}$$

Hence $D(t)\breve{w} = D(t)w = D(t)\breve{w}(u,t)$ and thus

$$D(t)w(u,t) = D(t)\breve{w}(u,t)\,.$$

□

We recall that for properly formulated DAEs, the inherent regular ODE is uniquely determined by the problem data whereas in [8], due to different projectors, a variety of inherent regular ODE had to be taken into account

Remark 2.6 *In particular, the matrices DA_1^{-1} and $-DA_1^{-1}BD^- = Dw'_u$ do not depend on the choice of P_0. Namely, with the two projectors $P_0, \check{P}_0$ along N_0, we have $A_1 = AD + BQ_0$ and $\check{A}_1 = AD + B\check{Q}_0$. We compute*

$$\check{A}_1 = AD + BQ_0\check{Q}_0 = A_1(P_0 + \check{Q}_0)$$

and hence

$$\check{A}_1^{-1} = (P_0 + \check{Q}_0)^{-1}A_1^{-1} = (\check{P}_0 + Q_0)A_1^{-1}\,,$$

obtaining $D\check{A}_1^{-1} = DA_1^{-1}$. Compute further

$$\begin{aligned} D\check{A}_1^{-1}B\check{D}^- &= D\check{A}_1^{-1}B\check{D}^- D\check{D}^- = D\check{A}_1^{-1}B\check{D}^- DD^- = D\check{A}_1^{-1}B\check{P}_0 D^- \\ &= D\check{A}_1^{-1}BD^- = DA_1^{-1}BD^-. \end{aligned}$$

We close this section with a solvability statement that follows the lines of [8].
We denote by $|||.|||_\infty$ the maximum norm on a compact interval $\mathcal{I}$.

Theorem 2.3 *Let the DAE (2.1) be tractable with index 1.*

(i) Through each $x_0 \in \mathcal{M}_0(t_0)$ passes exactly one solution of (2.1).

(ii) Given a solution $x_* \in C_D^1(\mathcal{I}, \mathbb{R}^m)$ of (2.1), $\mathcal{I}$ compact, $t_0 \in \mathcal{I}$, then all perturbed IVPs

$$f((Dx)'(t), x(t), t) = q(t), \quad D(t_0)(x(t_0) - x^0) = 0, \tag{2.9}$$

$x^0 \in \mathbb{R}^m, q \in C(\mathcal{I}, \mathbb{R}^m)$, are uniquely solvable on $C_D^1(\mathcal{I}, \mathbb{R}^m)$, supposed the perturbations $\|D(t_0)(x^0 - x_*(t_0))\|$ and $|||\, q \,|||_\infty$ are sufficiently small.

(iii) For the solution $x(.)$ of (2.9) it holds that

$$|||\, x - x_* \,|||_\infty \le K\{\|D(t_0)(x(t_0) - x_*(t_0))\| + |||\, q \,|||_\infty\}\,.$$

Proof.

(i) Using Lemma 2.1 we construct the inherent regular ODE (2.8). Then we solve the IVP for (2.8) with $u(t_0) = D(t_0)x_0$ and define the continuous function

$$x(t) := D(t)^- u(t) + Q_0(t)w(u(t), t), \quad x(t_0) = x_0.$$

Obviously, $D(t)x(t) = D(t)D(t)^- u(t) = R(t)u(t) = u(t)$ is continuously differentiable, thus $x(.) \in C_D^1$. From the ODE (2.8) we have

$$R(t)u'(t) = D(t)w(u(t), t)\,.$$

Therefore

$$\begin{aligned} 0 &= f(R(t)u'(t), D(t)^- u(t) + Q_0(t)w(u(t)t), t) \\ &= f((Dx)'(t), x(t), t). \end{aligned}$$

(ii) Denote $u_*(t) := D(t)x_*(t)$, $t \in \mathcal{I}$. By means of Lemma 2.4 below we provide a function $\tilde{w}(u, t, p)$ which is defined for

$$u \in U_\rho := \bigcup_{t \in \mathcal{I}} B(u_*(t), \rho), \;\; t \in \mathcal{I}, \;\; p \in B(0, \tau)$$

for certain $\rho > 0, \tau > 0$, such that

$$f(D(t)\tilde{w}, D(t)^- u + Q_0(t)\tilde{w}, t) - p = 0\,.$$

We consider $q(.) \in C(\mathcal{I}, I\!R^m)$, satisfying $|||\, q \,|||_\infty < \tau$, and $x^0 \in I\!R^m$, such that $\|D(t_0)(x^0 - x_*(t_0))\| < \rho$; thus if we denote $u^0 := D(t_0)x^0$, it holds that $\|u^0 - u_*(t_0)\| < \rho$. We solve the IVP

$$u'(t) - R'(t)u(t) = D(t)\tilde{w}(u(t), t, q(t)), \; u(t_0) = u^0.$$

Again, $u(t) = R(t)u(t)$ holds true, and

$$x(t) := D(t)^- u(t) + Q_0(t)\tilde{w}(u(t), t, q(t))$$

is the wanted solution.

(iii) From

$$\begin{aligned} u'(t) - u_*'(t) &= R'(t)(u(t) - u_*(t)) \\ &\quad + D(t)\left\{\tilde{w}(u(t), t, q(t)) - \tilde{w}(u_*(t), t, 0)\right\}, \qquad t \in \mathcal{I}, \\ u(t_0) - u_*(t_0) &= D(t_0)(x(t_0) - x_*(t_0)) \end{aligned}$$

we derive

$$\begin{aligned} u'(t) - u_*'(t) &= R'(t)\left(u(t) - u_*(t)\right) \\ &\quad + D(t) \int_0^1 \tilde{w}_u'(su(t) + (1-s)u_*(t), t, sq(t))\, ds \; (u(t) - u_*(t)) \\ &\quad + D(t) \int_0^1 \tilde{w}_q'(\cdots)\, ds \; q(t). \end{aligned}$$

Since $\mathcal{I}$ is compact, we may obtain uniform bounds for $D\tilde{w}_u' = -DA_1^{-1}BD, R'$ and $D\tilde{w}_p' = DA_1^{-1}$. In consequence,

$$\|u'(t) - u_*'(t)\| \leq L_1 \|u(t) - u_*(t)\| + L_2 \|q(t)\|,$$

what leads to

$$\|u(t) - u_*(t)\| \leq L_3 \left(\|D(t_0)(x(t_0) - x_*(t_0)\| + |||\, q \,|||_\infty\right).$$

Finally, in

$$\begin{aligned} x(t) - x_*(t) &= D(t)^- \left(u(t) - u_*(t)\right) \\ &\quad + Q_0(t) \int_0^1 \tilde{w}_u'(su(t) + (1-s)u_*(t), t, sq(t))\, ds \; (u(t) - u_*(t)) \\ &\quad + Q_0(t) \int_0^1 \tilde{w}_p'(\cdots)\, ds \; q(t) \end{aligned}$$

we may use bounds for $D^-, Q_0\tilde{w}_u' = -Q_0A_1^{-1}BD^-$ and $Q_0\tilde{w}_p' = Q_0A_1^{-1}$ to obtain the desired inequality. $\square$

As a simple consequence of the Implicit Function Theorem, analogously to Lemma 2.1, the following assertion results.

Lemma 2.4 *Let $t_0 \in \mathcal{I}$, $x_0 \in \mathcal{M}(t_0)$, $y_0 \in \operatorname{Im} D(t_0)$ such that*

$$N(t_0) \cap S(y_0, x_0, t_0) = \{0\}\,, \qquad f(y_0, x_0, t_0) = 0\,.$$

We denote

$$u_0 := D(t_0)x_0 \qquad w_0 := D(t_0)^- y_0 + Q(t_0)x_0\,,$$

and

$$\tilde{F}(w, u, t, p) := f(D(t)w, D(t)^- u + Q_0(t)w, t) - p$$

with $(w, u, t, p) \in I\!R^m \times I\!R^m \times I\!R \times I\!R^m$ in a neighbourhood of $(w_0, u_0, t_0, 0)$.
Then the equation $\tilde{F}(w, u, t, p) = 0$ determines implicitly the function $w = \tilde{w}(u, t, p)$ and

$$\begin{aligned}
\tilde{w}(u, t, p) &= \tilde{w}(R(t)u, t, p)\,, \\
\tilde{w}'_u &= -A_1^{-1} B D^-, \quad \tilde{w}'_p = A_1^{-1}\,, \\
\tilde{w}(u, t, 0) &= w(u, t)
\end{aligned}$$

with $w(u, t)$ the function from Lemma 2.1

3 Numerical integration

Once we have analyzed initial values problems for (2.2), in this section we study its numerical solution with Runge-Kutta methods and BDFs. Recall that for index-1 properly formulated DAEs, the leading term shows all the involved derivatives. For that reason, when we use an standard ODE method to solve the DAE (2.2), we can ensure that it is used only for the part of the solution which is derived and therefore we can expect good numerical results.

3.1 Numerical integration by Runge-Kutta methods

We consider an s-stages Runge-Kutta method with coefficients $(b^t, \mathcal{A})$. We assume that the matrix $\mathcal{A}$ is regular and the method is stiffly accurate, i.e. , $a_{si} = b_i$, $i = 1, \ldots, s$. We denote $t_{ni} = t_{n-1} + c_i h$, and $\alpha_{ij} = (\mathcal{A}^{-1})_{ij}$. As usual we assume that $\mathcal{A}1\!\!1 = c$ and thus $c_s = 1$. Given an approximation x_{n-1} of the solution of (2.2)at t_{n-1}, the new approximation x_n at $t_n = t_{n-1} + h$ is given by

$$x_n = X_{ns}$$

where the internal stages X_{ni}, $i = 1, \ldots, s$ are obtained solving the system

$$f([DX]'_{ni}, X_{ni}, t_{ni}) = 0, \qquad i = 1, \ldots, s \tag{3.1}$$

with the internal derivatives defined by

$$[DX]'_{ni} = \frac{1}{h} \sum_{j=1}^{s} \alpha_{ij} \left(D_{nj} X_{nj} - D_{n-1} x_{n-1} \right), \qquad i = 1, \ldots, s \, .$$

Observe that (3.1) is the numerical solution obtained when the equivalent problem

$$z - D(t)x = 0, \tag{3.2}$$
$$f((Rz)', x, t) = 0 \tag{3.3}$$

is solved with this RK method.

If we apply the decoupling process to (3.1), we obtain

$$R_{ni}[U]'_{ni} = D_{ni} w(U_{ni}, t_{ni}), \tag{3.4}$$
$$Q_{0,ni} X_{ni} = Q_{0,ni} w(U_{ni}, t_{ni}), \qquad i = 1, \ldots, s \, , \tag{3.5}$$

where $U_{ni} = D_{ni} X_{ni}$, $u_{n-1} = D_{n-1} x_{n-1}$ and

$$[U]'_{ni} = \frac{1}{h} \sum_{j=1}^{s} \alpha_{ij} \left(U_{nj} - u_{n-1} \right), \qquad i = 1, \ldots, s \, .$$

Thus the numerical solution can be decoupled into

$$x_n = D_n^- U_{ns} + Q_{0,n} X_{ns}$$

where U_{ns} is the last internal stage in (3.4) and $Q_{0,n} X_{ns}$ is given by (3.5).

Remark 3.1 *Observe that given x_{n-1}, we only advance with $D_{n-1} x_{n-1}$, and thus the errors in the nullspace of D are not propagated.*

Remark 3.2 *For non stiffly accurate RK methods, once we have computed the internal stages X_{ni}, we can define the numerical approximation at t_n as*

$$x_n = \rho x_{n-1} + (b^t \mathcal{A}^{-1} \otimes I_m) X_n$$

where $X_n = (X_{n1}, \ldots, X_{ns})$ and $\rho = 1 - b^t \mathcal{A}^{-1} \mathbb{1}$. But in this case, in general, $Q_{0,n} x_n$ does not satisfy equation (3.5), i.e. x_n may not belong to $\mathcal{M}_0(t_n)$.

In order to obtain a convergence result, we proceed like in the standard theory of numerical method for ODEs studying first the stability of the scheme. For this reason we consider the perturbed schemes

$$f([DX]'_{ni}, X_{ni}, t_{ni}) = \delta_{ni}, \qquad i = 1, \ldots, s\,, \tag{3.6}$$

where

$$[DX]'_{ni} = \frac{1}{h}\sum_{j=1}^{s} \alpha_{ij}\left(D_{nj}X_{nj} - D_{n-1}x_{n-1}\right), \qquad i = 1, \ldots, s\,,$$

and

$$f([D\bar{X}]'_{ni}, \bar{X}_{ni}, t_{ni}) = \bar{\delta}_{ni}, \qquad i = 1, \ldots, s\,, \tag{3.7}$$

where

$$[D\bar{X}]'_{ni} = \frac{1}{h}\sum_{j=1}^{s} \alpha_{ij}\left(D_{nj}\bar{X}_{nj} - D_{n-1}\bar{x}_{n-1}\right), \qquad i = 1, \ldots, s\,.$$

The decoupling process for (3.6) and (3.7) gives

$$\begin{aligned} R_{ni}[U]'_{ni} &= D_{ni}\tilde{w}(U_{ni}, t_{ni}, \delta_{ni}), \\ Q_{0,ni}X_{ni} &= Q_{0,ni}\tilde{w}(U_{ni}, t_{ni}, \delta_{ni}), \qquad i = 1, \ldots, s\,, \end{aligned} \tag{3.8}$$

and

$$\begin{aligned} R_{ni}[\bar{U}]'_{ni} &= D_{ni}\tilde{w}(\bar{U}_{ni}, t_{ni}, \bar{\delta}_{ni}), \\ Q_{0,ni}\bar{X}_{ni} &= Q_{0,ni}\tilde{w}(\bar{U}_{ni}, t_{ni}, \bar{\delta}_{ni}) \qquad i = 1, \ldots, s\,, \end{aligned} \tag{3.9}$$

respectively, where $\tilde{w}(u, t, p)$ is determined by Lemma 2.4. Recall that $\tilde{w}(u, t, 0) = w(u, t)$ with $w(u, t)$ the function obtained from Lemma 2.1.

An stable scheme is defined as follows.

Definition 3.1 *The Runge-Kutta method (3.1) is said to be stable if, for all sufficiently small perturbations $|\delta_{ni}| \le \tau$, $|\bar{\delta}_{ni}| \le \tau$, it holds that*

$$\|x_n - \bar{x}_n\| \le K\left(\|D_0x_0 - D_0\bar{x}_0\| + \max_{\ell \le n}\left\{\max_{1\le i\le s}\|\delta_{\ell i} - \bar{\delta}_{\ell i}\|\right\}\right), \qquad n \ge 1,$$

where K does not depend on the stepsizes used.

Theorem 3.1 *Let $x_* \in C^1_D([t_0, T], \mathbb{R}^m)$ be a solution of the index-1 tractable DAE (2.2). Let $\tilde{w}$ from Lemma 2.4 be given on $U \times [t_0, T] \times B(0, \tau)$, where $U \subseteq \mathbb{R}^m$ is a sufficiently large neighbourhood of the set $\{D(t)x_*(t) : t \in [t_0, T]\}$. Let $DA_1^{-1}BD^-$, A_1^{-1} and P_s be bounded. Then, the RK scheme applied to (2.2), with $t_n \le T$, is stable.*

Proof. We denote $E_{ni} = \bar{U}_{ni} - U_{ni}$, $e_{n-1} = u_{n-1} - \bar{u}_{n-1}$. Thus subtracting (3.8) from (3.9), we obtain

$$R_{ni}E'_{ni} = D_{ni}\left(\tilde{w}(\bar{U}_{ni}, t_{ni}, \bar{\delta}_{ni}) - \tilde{w}(U_{ni}, t_{ni}, \delta_{ni})\right) \tag{3.10}$$

where

$$E'_{ni} = \frac{1}{h}\sum_{j=1}^{s} \alpha_{ij}(E_{nj} - e_{n-1}), \qquad i = 1, \dots, s\,.$$

Observe that

$$R_{ni}E'_{ni} = E'_{ni} + \sum_{j=1}^{s} \alpha_{ij}\frac{1}{h}(R_{ni} - R_{nj})E_{nj} + \sum_{j=1}^{s} \alpha_{ij}\frac{1}{h}(R_{nj} - R_{n-1})e_{n-1}\,.$$

Thus denoting

$$\begin{aligned}
\tilde{W}_{ni} &= \int_0^1 \tilde{w}'_u(\tau\bar{U}_{ni} + (1-\tau)U_{ni}, t_{ni}, \delta_{ni})\, d\tau,\\
\tau_{ni} &= D_{ni}\left\{\tilde{w}(\bar{U}_{ni}, t_{ni}, \bar{\delta}_{ni}) - \tilde{w}(\bar{U}_{ni}, t_{ni}, \delta_{ni})\right\},\\
\phi_{ni} &= -\sum_{j=1}^{s} \alpha_{ij}\frac{1}{h}(R_{ni} - R_{nj})E_{nj} - \sum_{j=1}^{s} \alpha_{ij}\frac{1}{h}(R_{nj} - R_{n-1})e_{n-1},
\end{aligned}$$

we can write (3.10) as

$$E'_{ni} = D_{ni}\tilde{W}_{ni}E_{ni} + \tau_{ni} + \phi_{ni}, \qquad i = 1\dots, s\,. \tag{3.11}$$

With the notation $E'_n = (E'_{n1}, \dots, E'_{ns})$, and in a similar way E_n, τ_n and ϕ_n, and $\mathcal{D}_g = \text{diag}\,(g_{n1}, \dots, g_{ns})$, we rewrite (3.11) as

$$E'_n = \mathcal{D}_{D\tilde{W}}E_n + \tau_n + \phi_n\,,$$

and using that

$$hE'_n = (\mathcal{A}^{-1} \otimes I)\,(E_n - \mathbb{1} \otimes e_{n-1})\,,$$

we obtain

$$((\mathcal{A}^{-1} \otimes I) - h\mathcal{D}_{D\tilde{W}})E_n = (\mathcal{A}^{-1} \otimes I)(\mathbb{1} \otimes e_{n-1}) + h\tau_n + h\phi_n\,.$$

Now there exists an h_* such that for $h \le h_*$, the matrix $(I - h(\mathcal{A} \otimes I)\mathcal{D}_{D\tilde{W}})$ is nonsingular and

$$\|(I - h(\mathcal{A} \otimes I)\mathcal{D}_{D\tilde{W}})^{-1}\| \le 1 + h\,C_1$$

for some constant C_1 independent of the stepsize. Consequently

$$\begin{aligned}
E_n &= ((\mathcal{A}^{-1} \otimes I) - h\mathcal{D}_{D\tilde{W}})^{-1}\left\{(\mathcal{A}^{-1} \otimes I)(\mathbb{1} \otimes e_{n-1}) + h\tau_n + h\phi_n\right\}\\
&= (I - h(\mathcal{A} \otimes I)\mathcal{D}_{D\tilde{W}})^{-1}\left\{(\mathbb{1} \otimes e_n) + h(\mathcal{A} \otimes I)\tau_n + h(\mathcal{A} \otimes I)\phi_n\right\},
\end{aligned}$$

and hence

$$\|E_n\| \le (1 + hC_1)\|e_{n-1}\| + hC_2\|\tau_n\| + hC_2\|\phi_n\|\,. \tag{3.12}$$

As $R(t) \in C^1$, we have

$$\|\phi_n\| \le L\left(\|E_n\| + \|e_{n-1}\|\right)\,,$$

and thus for $h \le 1/(C_2 L)$, (3.12) gives

$$\|E_n\| \le (1 + hC_3)\|e_{n-1}\| + hC_2\|\tau_n\|\,,$$

and using that $e_n = E_{ns}$, we obtain

$$\|e_n\| \le (1 + hC_3)\|e_{n-1}\| + hC_2\|\tau_n\|\,.$$

At this point we simply have to observe that

$$\|\tau_{ni}\| \le \tilde{K}\|\delta_{ni} - \bar{\delta}_{ni}\|\,,$$

to get

$$\|e_n\| \le (1 + hC_3)\|e_{n-1}\| + hC_2\tilde{K}\|\delta_{ni} - \bar{\delta}_{ni}\|\,.$$

Now the standard recursion procedure gives the required stability bound for e_n,

$$\|e_n\| \le K_e\left(\|e_0\| + \max_{\ell \le n}\left\{\max_{1\le i\le s}\|\delta_{\ell i} - \bar{\delta}_{\ell i}\|\right\}\right)\,.$$

Finally, since

$$\begin{aligned}
x_n - \bar{x}_n &= D_n^- e_n + Q_{0,n}\left(\tilde{w}(u_n, t_n, \delta_{ns}) - \tilde{w}(\bar{u}_n, t_n, \bar{\delta}_{ns})\right) \\
&= D_n^- e_n + Q_{0,n}\left\{\int_0^1 \tilde{w}'_u(su_n + (1-s)\bar{u}_n, t_n, s\delta_{ns} + (1-s)\bar{\delta}_{ns})\,ds\right\}\ e_n \\
&\quad + Q_{0,n}\left\{\int_0^1 \tilde{w}'_p(\cdots)\,ds\right\}\ (\delta_{ns} - \bar{\delta}_{ns}) \\
&= \int_0^1 (I - Q_{0,n}A_1^{-1}(\cdots)B(\cdots))\,ds\ D_n^- e_n + \int_0^1 Q_{0,n}A_1^{-1}(\cdots)\,ds\ (\delta_{ns} - \bar{\delta}_{ns})
\end{aligned}$$

we may estimate

$$\|x_n - \bar{x}_n\| \le \hat{K}\|e_n\| + K_\delta\|\delta_{ns} - \bar{\delta}_{ns}\|$$

and obtain the required inequality.

□

From this stability result, now we can obtain the order of convergence for the RK method.

Theorem 3.2 *If the RK method satisfies the condition $C(q)$ and the solution $x(\cdot)$ of (2.2) satisfies $D(\cdot)x(\cdot) \in C^{q+1}$, then the method is convergent of order q.*

Proof. We consider the solution $x_*(t)$ in the perturbed scheme

$$f([Dx_*]'_{ni},\ x_*(t_{ni}), t_{ni}) = \tilde{\delta}_{ni}, \quad i = 1, \ldots, s$$

with

$$[Dx_*]'_{ni} = \frac{1}{h} \sum_{j=1}^{s} \alpha_{ij} (D_{nj} x_*(t_{nj}) - D_{n-1} x_*(t_{n-1})),$$

and the unperturbed scheme (3.1) for the numerical solution.
Due to stability,

$$\|x_n - x(t_n)\| \leq K \left(\|D_0 x_0 - D_0 x_*(t_0)\| + \max_{\ell \leq n} \max_{1 \leq i \leq s} \|\tilde{\delta}_{\ell i}\| \right) \tag{3.13}$$

As

$$\begin{aligned} \tilde{\delta}_{ni} &= A(x_*(t_{ni}), t_{ni})[Dx_*]'_{ni} - b(x_*(t_{ni}), t_{ni}) \\ &= A(x_*(t_{ni}), t_{ni}) \left\{ [Dx_*]'_n - (Dx_*)'(t_n) \right\}, \end{aligned}$$

the $C(q)$ condition implies that $\tilde{\delta}_{ni} = Q(h^q)$, and from (3.13) we obtain the desired result.

□

3.2 Numerical integration by BDFs

We consider now the numerical solution of properly formulated DAEs with k step variable coefficients BDFs, with $k \leq 6$. We proceed in a similar way to the study of Runge-Kutta methods in the previous subsection.
For BDFs, given an approximation x_{n-1} of the solution of (2.2) at t_{n-1}, the new approximation x_n at $t_n = t_{n-1} + h_n$ is obtained via

$$f([Dx]'_n, x_n, t_n) = 0 \tag{3.14}$$

where $[Dx]'_n$ is defined by

$$[Dx]'_n = \frac{1}{h_n} \sum_{j=0}^{k} \alpha_{nj} D_{n-j} x_{n-j} \, .$$

Observe that like for RK methods, (3.14) is the numerical solution obtained when the equivalent problem (3.2), (3.3) is solved with the BDF method.

We apply the decoupling process to (3.14) to obtain

$$R_n[u]'_n = D_n w(u_n, t_n) \tag{3.15}$$
$$Q_{0,n} x_n = Q_{0,n} w(u_n, t_n)\,, \tag{3.16}$$

where $u_n = D_n x_n$ and

$$[u]'_n = \frac{1}{h_n}\sum_{j=0}^{k} \alpha_{nj} u_{n-j}\,.$$

Thus the numerical solution can be written as

$$x_n = D_n^- u_n + Q_{0,n} x_n \in \mathcal{M}_0(t_n),$$

where u_n is obtained from in (3.15) and $Q_{0,n}x_n$ is given by (3.16). For BDFs, Remark 3.1 is also valid.

Again, to obtain a convergence result we proceed like in the standard theory of numerical methods for ODEs and firstly we study the stability of the scheme. That is why we consider the perturbed schemes

$$f([Dx]'_n, x_n, t_n) = \delta_n \tag{3.17}$$

where

$$[Dx]'_n = \frac{1}{h_n}\sum_{j=0}^{k} \alpha_{nj} D_{n-j} x_{n-j}\,,$$

and

$$f([D\bar{x}]'_n, \bar{x}_n, t_n) = \bar{\delta}_n \tag{3.18}$$

where

$$[D\bar{x}]'_n = \frac{1}{h_n}\sum_{j=0}^{k} \alpha_{nj} D_{n-j} \bar{x}_{n-j}\,.$$

The decoupling process for (3.17) and (3.18) gives

$$\begin{aligned} R_n[u]'_n &= D_n \tilde{w}(u_n, t_n, \delta_n) \\ Q_{0,n} x_n &= Q_{0,n} \tilde{w}(u_n, t_n, \delta_n)\,, \end{aligned} \tag{3.19}$$

and

$$\begin{aligned} R_n[\bar{u}]'_n &= D_n \tilde{w}(\bar{u}_n, t_n, \bar{\delta}_n) \\ Q_{0,n} \bar{x}_n &= Q_{0,n} \tilde{w}(\bar{u}_n, t_n, \bar{\delta}_n)\,. \end{aligned} \tag{3.20}$$

respectively.

An stable scheme is defined as follows.

Definition 3.2 *The BDF method (3.14) is called stable if for all perturbations* $\|\delta_n\| \le \tau,\ \|\tilde{\delta}_n\| \le \tau$, τ *sufficiently small, it holds that*

$$\|x_n - \bar{x}_n\| \le K \left(\max_{0\le \ell \le k-1} \|D_\ell x_\ell - D_\ell \bar{x}_\ell\| + \max_{n\ge k} \|\delta_n - \bar{\delta}_n\| \right) .$$

Recall that in the regular ODE case there are restrictions to choose the stepsize for the variable coefficient BDF. Thus we only consider grids $t_0 < t_1 < \cdots < t_n, t_n \le T$, such that there is a $I\!R^k$-norm $\|\cdot\|_*$ with $\|\mathcal{F}_n\|_* \le 1$ for $n \ge k$ and all grids, where

$$\mathcal{F}_n = \begin{pmatrix} -\tilde{\alpha}_{n_1} & & & -\tilde{\alpha}_{n_k} \\ 1 & & & \\ & \ddots & & \\ & & 1 & 0 \end{pmatrix}, \qquad \tilde{\alpha}_{ni} := \alpha_{ni}/\alpha_{n_0} .$$

Observe that in the constant stepsize case, $\mathcal{F}_n = \mathcal{F}$ is constant and $\|\cdot\|_*$ exists due to Dahlquist's root criterion for $k \le 6$.

Theorem 3.3 *Let the assumptions of Theorem 3.1 be given. Then the BDF applied to (2.2) is stable on grids with* $\|\mathcal{F}_n\|_* \le 1$.

Proof. Subtracting (3.19) from (3.20) and denoting $e_{n-1} = \bar{u}_{n-1} - u_{n-1}$ we obtain

$$R_n e_n' = D_n \left(\tilde{w}(\bar{u}_n, t_n, \bar{\delta}_n) - \tilde{w}(u_n, t_n, \delta_n) \right) \tag{3.21}$$

where

$$e_n' = \frac{1}{h_n} \sum_{j=0}^{k} \alpha_{nj} e_{n-j} .$$

Observe that

$$R_n e_n' = e_n' + \sum_{j=0}^{k} \alpha_{nj} \frac{1}{h_n} (R_n - R_{n-j}) e_{n-j} .$$

Thus denoting $\tilde{\alpha}_{nj} = \alpha_{nj}/\alpha_{n0}$ and

$$\begin{aligned} \tilde{W}_n &= \int_0^1 \tilde{w}_u'(\tau \bar{u}_n + (1-\tau) u_n, t_n, \delta_n)\, d\tau , \\ \tau_n &= D_n \left\{ \tilde{w}(\bar{u}_n, t_n, \bar{\delta}_n) - \tilde{w}(\bar{u}_n, t_n, \delta_n) \right\} , \\ \phi_n &= \sum_{j=1}^{k} \tilde{\alpha}_{nj} \frac{1}{h_n} (R_n - R_{n-j}) e_{n-j} , \end{aligned}$$

we can write (3.21) as

$$e_n = -\sum_{j=1}^{k} \tilde{\alpha}_{nj} e_{n-j} + \frac{h_n}{\alpha_{n0}} D_n \tilde{W}_n e_n + \frac{h_n}{\alpha_{n0}} \tau_n - h_n \phi_n . \tag{3.22}$$

Now there exists an h_* such that for $h_n \leq h_*$, the matrix $(I - \frac{h_n}{\alpha_{n0}} D_n \tilde{W}_n)$ is nonsingular and

$$\left\| \left(I - \frac{h_n}{\alpha_{n0}} D_n \tilde{W}_n \right)^{-1} \right\| \leq 1 + hC \, .$$

for some constant independent of the stepsize. Denoting $H_n := (I - \frac{h_n}{\alpha_{n0}} D_n \tilde{W}_n)$, equation (3.22) can be written as

$$e_n = H_n^{-1} \left(- \sum_{j=1}^{k} \tilde{\alpha}_{nj} e_{n-j} + \frac{h_n}{\alpha_{n0}} \tau_n - h_n \phi_n \right) . \tag{3.23}$$

With the notation

$$\begin{aligned} E_n &= (e_n, \ldots, e_{n-k+1}) \, , & T_n &= (H_n^{-1} \tau_n, \, 0, \ldots, 0) \, , \\ \Psi_n &= (D_n \tilde{W}_n H_n^{-1} \textstyle\sum_{j=1}^{k} \alpha_{nj} e_{n-j}, 0, \ldots, 0) \, , & \Phi_n &= (\alpha_{n0} H_n^{-1} \phi_n, 0, \ldots, 0) \, , \end{aligned}$$

we rewrite (3.23) as

$$E_n = (\mathcal{F}_n \otimes I) E_{n-1} + \frac{h_n}{\alpha_{n0}} T_n - \frac{h_n}{\alpha_{n0}} \Phi_n - \frac{h_n}{\alpha_{n0}} \Psi_n \, .$$

Hence taking into consideration that there is a norm such that $\|\mathcal{F}_n\|_* \leq 1$ for all n, we have,

$$\|E_n\|_* \leq \|E_{n-1}\|_* + h a_n \|T_n\|_* + h a_n \|\Psi_n + \Phi_n\|_* \, ,$$

where $a_n = 1/|\alpha_{n0}|$. Now, as $\|T_n\|_* \leq k\|\delta_n - \bar{\delta}_n\|$ and $\|\Phi_n + \Phi_n\|_* \leq L\|E_{n-1}\|_*$, using the standard recursion procedure we obtain the required stability bound for e_n. Proceeding with $x_n - \bar{x}_n$ like in the RK case, we obtain the stability bound.

□

Theorem 3.4 *We consider the k-step BDF. If the solution $x_*(\cdot)$ of (2.2) satisfies $D(\cdot)x_*(\cdot) \in C^{k+1}$ and $D_\ell x_\ell - D_\ell x_*(t_\ell) = O(h^k), \ell = 0, \ldots, k-1$, then the method is convergent with order k.*

Proof. We consider the solution $x_*(t)$ in the perturbed scheme

$$f([Dx_*]'_n, \, x_*(t_n), t_n) = \tilde{\delta}_n,$$

with

$$[Dx_*]'_n = \frac{1}{h_n} \sum_{j=0}^{k} \alpha_{nj} \, x_*(t_{n-j}),$$

and the unperturbed scheme (3.14) for the numerical solution.
Due to stability,

$$\|x_n - x_*(t_n)\| \le K \left(\max_{0\le\ell\le k-1} \|D_\ell x_\ell - D_\ell x_*(t_\ell)\| + \max_{n\le k} \|\tilde{\delta}_n\| \right) \tag{3.24}$$

As

$$\begin{aligned}\tilde{\delta}_n &= A(x_*(t_n), t_n)[Dx_*]'_n - b(x_*(t_n), t_n) \\ &= A(x_*(t_n), t_n)\{[Dx_*]'_n - (Dx_*)'(t_n)\}\,,\end{aligned}$$

the consistency with order k of the BDF implies $\tilde{\delta}_n = O(h^k)$. From (3.24) we obtain the desired result.

□

As expected, the standard integration methods work well on compact intervals. Choosing sufficiently fine grids we obtain sufficiently close approximation. This applies also to the special DAE (1.2) in both its well formulated versions (1.3) and (1.6).
In the following section we deal with the qualitative behaviour of the numerical approximations on infinite intervals. In particular, we would like to point out the difference of (1.2) and (1.6) for example (1.1).

4 Numerically well formulated DAEs

As we have seen in Section 2, for the properly formulated index-1 DAE (2.1), the inherent regular ODE is

$$(Dx)' = R'Dx + w(Dx, t)$$

and has $\operatorname{Im} D(t)$ as invariant subspace. Let's see the consequences of the condition $\operatorname{Im} D(t)$ to be constant on the inherent regular ODE.

Lemma 4.1 *For the index 1 case, if* $\operatorname{Im} D(t)$ *is constant, then the solution component* Dx *satisfies*

$$(Dx)' = w(Dx, t). \tag{4.1}$$

Proof. For the index 1 case, in the inherent regular ODE we have the terms $R'(t)D(t)x(t)$, with $\operatorname{Im} R(t) = \operatorname{Im} D(t)$. Thus if $\operatorname{Im} D(t)$ is constant, then we can find a constant projector V such that $\operatorname{Im} V = \operatorname{Im} D(t)$. As $R(t)$ is a projector and $\operatorname{Im} V = \operatorname{Im} R(t)$, we have $R(t)V = V$ and hence also $R'(t)V = 0$. On the other hand, as V is a projector and $\operatorname{Im} V = \operatorname{Im} D(t)$, we have $VD(t) = D(t)$ and thus $(I-V)D(t) = 0$.Therefore

$$R'(t)D(t) = R'(t)VD(t) + R'(t)(I-V)D(t) = 0\,.$$

□

The condition $\operatorname{Im} D(t)$ to be constant also has nice consequences for the numerical solution.

Theorem 4.2 *We consider the numerical solution of an index-1 DAE with an stiffly accurate RK method or a BDF method. Assume that* $\operatorname{Im} D(t)$ *is constant. Then the numerical solution can be decoupled into*

$$x_n = D_n^- D_n x_n + Q_{0,n} x_n$$

where

1. $D_n x_n$ *is the numerical solution with the RK method or the BDF applied to the regular ODE (4.1).*
2. $Q_{0,n} x_n = Q_{0,n} w(x_n, t_n)$.

Proof. First, observe that from the definition of $[DX]'_{ni}$ for RK methods or $[Dx]'_n$ for BDF, for any constant matrix M we have

$$M[DX]'_{ni} = [MDX]'_{ni}, \qquad M[Dx]'_n = [MDx]'_n\,.$$

Applying to the RK method (3.1) a similar decoupling process to the one used to obtain the inherent regular ODE, we obtained (3.4)-(3.5). As $\operatorname{Im} D(t)$ is constant, there exists a constant projector V such that $\operatorname{Im} D(t) = \operatorname{Im} V$. For such a projector we have $R(t)V = V$, $VD(t) = D(t)$, $(I-V)D(t) = 0$. Therefore

$$R_{ni}[DX]'_{ni} = R_{ni}V[DX]'_{ni} + R_{ni}(I-V)[DX]'_{ni} = V[DX]'_{ni} = [DX]'_{ni}$$

It means that we are solving the inherent regular ODE (4.1) with the RK method.

In a similar way, for the BDF (3.14), the decoupling process gives (3.15)-(3.16). Proceeding in a similar way as it is done for RK methods we obtain $R_n[Dx]'_n = [Dx]'_n$ which implies that we are also integrating the inherent regular ODE (4.1) with the BDF method. □

Hence if the inherent regular ODE is contractive for a certain norm, then the approximations $D_n x_n$, given by an algebraically stable RK method behave also contractively for that norm.

The fact that $\operatorname{Im} D(t)$ constant implies that we are actually integrating numerically the inherent regular ODE on the right subspace, allow us to give the following definition.

Definition 4.1 *We will say that index-1 DAEs with constant* $\operatorname{Im} D(t)$ *are* **numerically well formulated**.

With other words, numerically well formulated DAEs are those, where the discretization and the decoupling commute.
As we could realize in Example 1.1, if the DAE is not numerically well formulated, for algebraically stable RK methods we may have strong restrictions on the stepsize h in order to reflect the asymptotical behavior of the solution.

With the concept of numerically well formulated DAEs, we can explain within the same framework the known conditions that ensure a good qualitative behaviour for the numerical solution, namely, the subspace $\operatorname{Ker} A(x,t)$ to be constant, the subspace $S(t)$ to be constant, and the formulation (1.4) with constant subspace $\operatorname{Im} A(t)$.

Constant subspace $\mathbf{Ker A(x,t)}$. Recall from [8] that, if $\operatorname{Ker} A(x,t)$ is constant the equation

$$A(x,t)x' + b(x,t) = 0 \tag{4.2}$$

should be rewritten as

$$A(x,t)(Px)' + b(x,t) = 0 \tag{4.3}$$

using a constant projector P along $\operatorname{Ker} A(x,t)$. Here we have $D := P,\ R := P$, thus a numerically well formulated DAE results, just with constant R. This confirms the positive results in [8] on contractivity, etc. In particular we have:

Proposition 4.3 *Given the index-1 DAE (4.2) and its numerical solution with an stiffly accurate RK method or a BDF. If* $\operatorname{Ker} A(x,t)$ *is constant, then (4.2) is numerically equivalent to the numerically well formulated DAE (4.3).*

Constant subspace $\mathbf{S(t)}$. We consider the index-1 homogeneous equation (1.2), i.e.

$$A(t)x'(t) + B(t)x(t) = 0\,. \tag{4.4}$$

As for any smooth projector $P(t)$ along $\operatorname{Ker} A(t)$ it holds that $PP'P = 0$ and for the homogeneous problem we have $x(t) = P_s(t)x(t)$, with $P_s(t)$ the canonical projector, we obtain that if $P_s(t)$ is differentiable, (4.4) is equivalent to the equation

$$A(t)(P_s(t)x(t))' + B(t)x(t) = 0\,, \tag{4.5}$$

which has a properly formulated leading term. It turns out that problems (4.4) and (4.5) are also numerically equivalent when they are solved with stiffly accurate RK method or BDFs. To see this observe that the numerical solution of (4.4) with a stiffly accurate RK method satisfies $X_{ni} = P_{s,ni}X_{ni}$ and thus, instead of

$$A_{ni}X'_{ni} + B_{ni}X_{ni} = 0\,, \qquad i = 1, \ldots, s\,,$$

with

$$X_{ni} = x_{n-1} + h\sum_{j=1}^{s} a_{ij}X'_{nj}\,, \qquad i = 1, \ldots, s\,,$$

we can write equivalently

$$A_{ni}[P_sX]'_{ni} + B_{ni}P_{s,ni}X_{ni} = 0\,, \qquad i = 1, \ldots, s\,,$$

with

$$P_{s,ni}X_{ni} = P_{s,n-1}x_{n-1} + h\sum_{j=1}^{s} a_{ij}[P_sX]'_{nj}\,, \qquad i = 1, \ldots, s\,,$$

that means that we are actually integrating (4.5). Similar reasoning gives the result for BDFs.
Now, if $\operatorname{Im} P_s(t) = S(t)$ is constant, then the DAE (4.4) is numerically well formulated. We summarize the above lines in the next Proposition.

Proposition 4.4 *Given the linear, homogeneous index-1 DAE (4.4) and its numerical solution with an stiffly accurate RK method or a BDF. If $S(t)$ is constant, then (4.4) is numerically equivalent to the numerically well formulated DAE (4.5).*

As the following example shows, if $\operatorname{Im} P_s(t) = S(t)$ is not constant, then we may have restrictions on the stepsize h in fact.

Example 4.1 *Recall once more example (1.1). For this problem*

$$S(t) = \left\{ x \,:\, x_1 = \frac{\delta t - 1}{1-\delta}x_2 \right\} \qquad \operatorname{Ker} A(t) = \left\{ x \,:\, x_1 = \frac{\delta t}{1-\delta}x_2 \right\}.$$

And thus $\operatorname{Im} P_s(t)$ is not constant. Remember that in this case for the implicit Euler method we have strong stepsize restriction.

On the other hand, as the following example stresses, the fact that $\operatorname{Im} P_s(t)$ is not constant does not necessarily imply stepsize restrictions.

Example 4.2 *We consider the DAE*

$$\begin{pmatrix} 0 & \alpha t \\ 0 & 1 \end{pmatrix} x' + \begin{pmatrix} -t & \alpha t(\alpha + t) \\ 1 & \alpha(1-t) \end{pmatrix} x = 0 \qquad x_2(0) = 1$$

which has index 1 for $1 - \alpha t \neq 0$ *and whose solution is*

$$x_1(t) = \alpha\, t\, x_2(t)\,, \qquad x_2(t) = e^{-\alpha t} x_{2,0}\,.$$

In this case $S(t) = \{x \in \mathbb{R}^2 \,:\, x_1 = \alpha t x_2\}$*, varies with time, and thus apparently (4.5) is not numerically well formulated. Nevertheless, the numerical solution with the implicit Euler method is*

$$x_{1,n} = \alpha t_n\, x_{2,n}\,, \qquad x_{2,n} = \frac{1}{1+\alpha h} x_{1,n-1}\,.$$

What happens in this case is that Ker $A(t) = \{x \in \mathbb{R}^2 : x_2 = 0\}$ *is constant and the problem is numerically equivalent to*

$$A(t)(Px)' + B(t)x = 0$$

with P *a constant projector along Ker* $A(t)$*, which is numerically well formulated (cf, Proposition 4.3).*

Formulation (1.4) with constant subspace ImA(t). Modified BDF methods and modified RK method to integrate linear variable coefficients DAEs have been proposed in [4], [6] respectively. Both approaches are based on the numerical integration of the DAE (1.4), i.e.

$$(A(t)x)' + (B(t) - A'(t))x = 0\,. \tag{4.6}$$

This DAE is not properly formulated but the equivalent DAE

$$R_A(t)(A(t)x)' - R_A(t)A(t)'x(t) + B(t)x(t) = 0 \tag{4.7}$$

where $R_A(t)$ is a projector onto $\operatorname{Im} A(t)$, has well matched coefficients R_A and A.

In [6], [7] contractivity conditions for the $A(t)x(t)$ part of the solution were studied. It was proved that for stiffly accurate algebraically stable RK methods, , provided that $\operatorname{Im} A(t)$ is constant, the $A(t_{n+1})x_{n+1}$ part of the numerical solution for (4.6) is also contractive.

We can now give an explanation of this fact.

Proposition 4.5 *Given the index-1 DAE (4.6) and its numerical solution with an stiffly accurate RK method or a BDF. If* $\operatorname{Im} A(t)$ *is constant, and* R_A *denotes a constant projector onto* $\operatorname{Im} A(t)$*, then (4.6) is numerically equivalent to the numerically well formulated DAE*

$$R_A(A(t)x)' - R_A A(t)'x(t) + B(t)x(t) = 0\ . \tag{4.8}$$

The example (1.1) has $\operatorname{Im} A(t)$ constant thus the DAE formulated as (4.8) is numerically well formulated. The inherent regular ODE constructed with Ax is contractive for $\delta - \mu > 0$ and thus the Ax part of the numerical solution also has this contractivity behavior for any stepsize. When this equation was integrated in [6], [7] as (4.6), we were actually integrating the numerically well formulated DAE (4.8).

5 The general index-1 case

Sometimes more general descriptions as (2.1) arise, namely

$$A(x(t),t)(d(x(t),t))' + b(x(t),t) = 0, \tag{5.1}$$

where $d : \mathcal{D}_0 \times \mathcal{I}_0 \to I\!R^m$ is an additional continuous function, that has a continuous partial derivative $d'_x(x,t) =: D(x,t)$.

Definition 5.1 *Equation (5.1) has a properly formulated leading term if*

$$\operatorname{Ker} A(x,t) \oplus \operatorname{Im} D(x,t) = I\!R^m \qquad \textit{for all } x \in \mathcal{D}_0, t \in \mathcal{I}_0\,,$$

and there is a projector function $R \in C^1(\mathcal{I}_0, L(I\!R^m))$ *such that* $R(t)^2 = R(t)$, $\operatorname{Ker} A(x,t) = \operatorname{Ker} R(t)$, $\operatorname{Im} D(x,t) = \operatorname{Im} R(t)$ *and* $d(x,t) = R(t)d(x,t)$ *for all* $x \in \mathcal{D}_0, t \in \mathcal{I}_0$.

In particular, $\operatorname{Ker} A(x,t)$ and $\operatorname{Im} D(x,t)$ do not depend on x , i.e. one should try arrange things in such a way. Note that, if at the beginning only one of these subspaces is independent of x, we may have both of them independent of x in a new version by simply rearrangements.

If (5.1) has a properly formulated leading term, the enlarged system

$$\left.\begin{array}{rcl} A(x(t),t)(R(t)y(t))' + b(x(t),t) & = & 0 \\ y(t) - d(x(t),t) & = & 0 \end{array}\right\} \tag{5.2}$$

represents a properly formulated DAE of type (2.1).
If $x_* \in C$ with $d(x_*(.),.) \in C^1$ solves the original DAE (5.1), then the pair $(x_*, y_*) \in C$, $y_*(.) := d(x_*(\cdot),\cdot)$, $R(\cdot)y_*(\cdot) = y_*(\cdot) \in C^1$ satisfies (5.2) and vice versa. In this sense are (5.1) and (5.2) equivalent.

Definition 5.2 *The DAE (5.1) with properly formulated leading term is said to be tractable with index 1, if*

$$\operatorname{Ker} A(x,t) \cap S_0(y,x,t) = \{0\} \quad \textit{for } x \in \mathcal{D}_0\,,\ t \in \mathcal{I}_0\,,\ y \in I\!R^m\ .$$

Theorem 5.1 *The DAE (5.1) and its enlarged system (5.2) are index 1 tractable at the same time.*

Proof. We put (5.2) in the form $\hat{A}(\hat{x},t)(\hat{D}\hat{x})' + \hat{b}(\hat{x},t) = 0, \hat{x} = \begin{pmatrix} y \\ x \end{pmatrix}$ and compute

$$\hat{N}_0(t) \cap \hat{S}_0(\hat{y},\hat{x},t) = \left\{ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \in I\!R^m \times I\!R^m : z_1 = D(x,t)z_2, \right.$$
$$\left. z_2 \in N_0(x,t) \cap S_0(y,x,t) \right\} .$$

□

Remark 5.1 *Via Theorem 2.3, solvability statements are given for (5.2), hence also for (5.1).*

In Section 3 we have mentioned the equivalence of the RK methods and BDF applied to (2.1) and its enlarged system (3.2), (3.3). The same holds true for (5.1) and (5.2). Therefore, the above assertions about stability and convergence of RK methods and BDFs can be used to obtain stability and convergence results for these methods when they are applied to (5.1). In the consequence, (5.2) and so also (5.1) are **numerically well formulated** supposed $\operatorname{Im} R(t) = \operatorname{Im} D(x,t)$ does not vary at all.

One could think writing (5.1) better as

$$A(x(t),t)d_x'(x(t),t)x'(t) + b(x(t),t) + A(x(t),t)d_t'(x(t),t) = 0 \ . \tag{5.3}$$

However, for singular $A(x,t)d_x'(x,t)$ there is again the problem on the precise meaning of the leading term. In this sense applying numerical integration methods to the DAE (5.3) is not only (sometimes much) more expensive but also mathematically somewhat wrong, except for the case of a constant nullspace $\operatorname{Ker}(A(x,t)d_x'(x,t))$.

6 Conclusions

When modeling complex processes by means of DAEs, at the beginning, one should carefully investigate the leading term and propose properly formulated problems. From the numerical point of view, working with well formulated problems means in some sense that the ODE method is used to discretize only the differential components. The discussion of numerical integration methods has made clear that the case of a time invariant $\operatorname{Im} D(t)$, i.e. numerically well formulated problems, is essentially favourable. If it is possible, one should use a numerically well formulated version. On a different place we will discuss how to realize this property when modeling problems.

Let us conclude this paper mentioning that there is some hope to handle also higher index DAEs better.
The famous index-2 example treated in [13]

$$\begin{pmatrix} 0 & 0 \\ 1 & \eta t \end{pmatrix} x' + \begin{pmatrix} 1 & \eta t \\ 0 & 1+\eta \end{pmatrix} x = \begin{pmatrix} g(t) \\ 0 \end{pmatrix}, \qquad \eta \neq -1,$$

is known to bring e.g. the implicit Euler method in big trouble. However, applying the same method to a numerically well formulated version, i.e. (1.6) , it works fine.

References

[1] K. Balla, R. März: An unified approach to linear differential algebraic equations and their adjoint equations. Preprint 2000-18, Humboldt-Universität Berlin, Institut für Mathematik.

[2] K. Balla, R. März: Linear differential algebraic equations of index 1 and their adjoint equations. Results in Math. 37 (2000) 13-35.

[3] K. E. Brenan, S. L. Campbell, L. R. Petzold: Numerical solution of initial value problems in differential-algebraic equations. North Holland, Amsterdam, 1989.

[4] K. D. Clark: Difference methods for the numerical solution of time varying singular systems of differential equations. SIAM J. alg. disc. math. 7(2), 1986, 236-246.

[5] H. Gajewski, K. Gröger, K. Zacharias: Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen. Akademie-Verlag Berlin, 1974.

[6] B. García Celayeta: Stability for differential algebraic equations. PhD-Thesis, Universidad Pública de Navarra, Pamplona, 1998.

[7] B. García Celayeta, I. Higueras: Runge-Kutta methods for DAEs. A new approach. J. Computational and Applied Mathematics 111 (1-2), 1999, 49-61.

[8] E. Griepentrog, R. März: Differential-algebraic equations and their numerical treatment. Teubner, Leipzig 1986.

[9] E. Hairer, G. Wanner: Solving ordinary differential equations II. Stiff and differential-algebraic problem. Springer Berlin Heidelberg, 1991.

[10] M. Hanke, R. März: On asymptotics in case of daes. ZAMM 76 (1996) Suppl. 1, 99-102.

[11] R. März: Eine einheitliche Sicht auf lineare Algebro-Differentialgleichungen und ihre adjungierten Systeme. Proc. 5. Workshop über Deskriptorsysteme, Paderborn, 11. - 15. März 2000.

[12] R. März: On initial value problems in differential-algebraic equations and their numerical treatment. Computing 35, 1985, 13-37.

[13] L. R. Petzold: Differential/algebraic equations are not ODEs. SIAM J. Sci. Statict Comput. 3 (1982), 367-384.

[14] A. M. Stuart, A. R. Humphries: Dynamical systems and numerical analysis. Cambridge University Press, Cambridge, UK, 1998.